

---

# **IDENTIFICATION AND CHARACTERIZATION OF GLYCOSIDE HYDROLASES FOR BIOTECHNOLOGICAL APPLICATIONS**

---

**Andrea Strazzulli**

Dottorato in Scienze Biotecnologiche – XXIII ciclo  
Indirizzo Biotecnologie Industriale e Molecolare  
Università di Napoli Federico II





Dottorato in Scienze Biotecnologiche –XXIII ciclo  
Indirizzo Biotecnologie Industriale e Molecolare  
Università di Napoli Federico II



---

# **IDENTIFICATION AND CHARACTERIZATION OF GLYCOSIDE HYDROLASES FOR BIOTECHNOLOGICAL APPLICATIONS**

---

**Andrea Strazzulli**

Dottorando:	Andrea Strazzulli
Relatore:	Prof. Mosè Rossi
Correlatore:	Dott. Marco Moracci
Coordinatore:	Prof. Giovanni Sannia





*Ai  
miei genitori*



## Index

<i>Riassunto</i>	8
<i>Abbreviations</i>	18
<b>General introduction</b>	<b>19</b>
<i>The Glycoside Hydrolases</i>	20
<i>The CAZy Database</i>	20
<i>Biotechnological application of Glycoside Hydrolases</i>	21
<i>Catalytic mechanisms of Glycoside Hydrolases</i>	23
Identification of the catalytic residues of glycoside hydrolases	25
The glycosynthases approach	27
<b>Aim of the thesis</b>	<b>31</b>
<b>Experimental procedures</b>	<b>33</b>
<i>Bacterial Strains</i>	34
<i>Culture media</i>	34
<i>Chemicals</i>	34
<i>Detection of transgalactosylation activity by Thin Layer Chromatography (TLC)</i>	34
<i>Site-directed mutagenesis</i>	34
<i>Characterization of Aa<math>\beta</math>-gal wild type and mutants</i>	35
Expression and purification of recombinant Aa $\beta$ -gal wild type and mutants	35
Enzymatic characterization of Aa $\beta$ -gal wild type and mutants	36
Nano-HPLC-ESI-MS/MS Experiments	37
<i>Characterization of TmGalA wild type and mutant</i>	37
Expression and purification of TmGalA wild type and Asp327Gly mutant	37
Enzymatic characterization of TmGalA Asp327Gly	38
Transgalactosylation trials of TmGalA Asp327Gly	38
Characterization of the galactosylated oligosaccharides obtained by TmGalA Asp327Gly	38
<i>Identification of the catalytic residues of SS01353</i>	39
Expression and purification of SS01353 wild type and mutants	39
Standard assay of SS01353 and mutants	40
Chemical rescue activity of SS01353 D462G mutant	40
Analysis of the products obtained by chemical rescue	40
Inhibition of SS01353 wild type	40
Nano-ESI-MS of Intact Protein Samples	40
Nano-HPLC-ESI-MS/MS Experiments	41
<i>Characterization of the substrate specificity of nSs<math>\alpha</math>-man on mannosylated glycans</i>	42
Purification of native $\alpha$ -mannosidase from <i>S. solfataricus</i> (nSs $\alpha$ -man)	42
Expression and purification of rSs $\alpha$ -man	42
Enzymatic assays	43
Reactivation of the mutant Ss $\alpha$ -man Asp338Gly	43
Activity of Ss $\alpha$ -man on N-linked oligosaccharides	43
De-mannosylation analysis of Rnase B	43
<b>Chapter I: A <math>\beta</math>-galactosidase from <i>Alicyclobacillus acidocaldarius</i></b>	<b>45</b>
<b>Introduction</b>	<b>45</b>
<i>The <math>\beta</math>-Galactosidases</i>	46
The $\beta$ -galactosidases of GH42 family	47
Biotechnological applications of $\beta$ -galactosidases	47
The $\beta$ -galactosidase from <i>A. acidocaldarius</i>	48
<b>Results and Discussion</b>	<b>50</b>

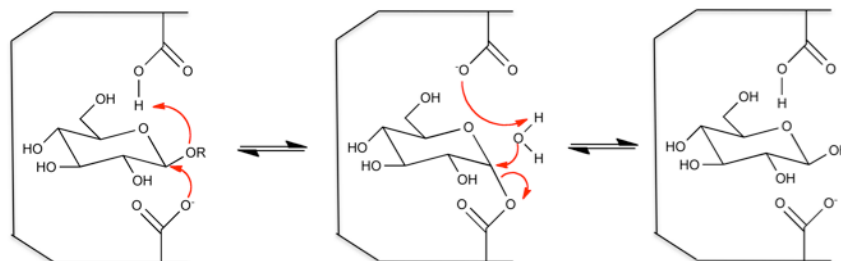
<i>Characterization of <math>\beta</math>-galactosidase mutants from A. acidocaldarius for the oligosaccharides synthesis</i>	51
Identification of catalytically active residues in Aa $\beta$ -gal	51
3D-structure prediction of Aa $\beta$ -gal	52
Site-directed mutagenesis, expression and purification of Aa $\beta$ -gal mutants	54
Analysis of steady-state kinetic constants of Aa $\beta$ -gal mutants	54
Chemical rescue of Aa $\beta$ -gal Mutants	55
Transglycosylation activity of Aa $\beta$ -gal mutant Glu361Gly	58
Analysis of $\beta$ -galactosynthet activity of mutant Aa $\beta$ -gal Glu313Gly	62
<b>Chapter II: A <math>\alpha</math>-galactosidase from <i>Thermotoga maritima</i></b>	<b>70</b>
<b>Introduction</b>	<b>70</b>
<i>The <math>\alpha</math>-galactosidases</i>	71
Applications of $\alpha$ -galactosidases	71
<b>Results and Discussion</b>	<b>73</b>
<i>Kinetic characterization and chemical rescue of TmGalA nucleophile mutants.</i>	74
<i>Characterization of the <math>\alpha</math>-galactosynthet activity of mutant TmGalA Asp327Gly</i>	75
$\alpha$ -galacto-oligosaccharide synthesis by the mutant Asp327Gly.	75
<i>Reaction mechanism of <math>\alpha</math>-D-galactosynthase</i>	78
<b>Chapter III: A <math>\beta</math>-glycosidase from <i>Sulfolobus solfataricus</i></b>	<b>80</b>
<b>Introduction</b>	<b>80</b>
<i>A new <math>\beta</math>-glycosidase from S. solfataricus</i>	81
<b>Results</b>	<b>82</b>
<i>Isolation of ORF SS01353</i>	83
<i>Identification of the catalytic residues of SS01353</i>	85
Expression and purification of SS01353 wild type and mutants	86
Reactivation of SS01353 mutants	87
Identification of the catalytic nucleophile of SS01353	88
<b>Discussion</b>	<b>95</b>
<b>Chapter IV: A <math>\alpha</math>-mannosidase from <i>Sulfolobus solfataricus</i></b>	<b>99</b>
<b>Introduction</b>	<b>99</b>
N-glycosylation in Archaea	100
Classification of $\alpha$ -mannosidases	101
The Ss $\alpha$ -man from <i>S. solfataricus</i>	102
<b>Results</b>	<b>103</b>
<i>Characterization of Ss<math>\alpha</math>-man</i>	104
Identification of the catalytic residues of Ss $\alpha$ -man	104
Activity of Ss $\alpha$ -man on high-mannose N-linked oligosaccharides	105
Activity of Ss $\alpha$ -Man on Ribonuclease B glycoprotein	107
<b>Discussion</b>	<b>109</b>
<b>References</b>	<b>112</b>
<i>List of Publications</i>	122
<i>Meetings Communications</i>	122
<i>Foreign experience.</i>	123

## Riassunto

La sintesi e la degradazione di carboidrati, sotto forma di di-, oligo-, e polisaccaridi, sono fenomeni importantissimi in tutti e tre i domini dei viventi. Queste molecole, che rappresentano i biopolimeri più abbondanti in natura, svolgono infatti funzioni di vitale importanza come fonti di energia, elementi strutturali e di regolazione intra- ed intercellulare.

La grande diversità strutturale dei carboidrati riflette un'altrettanto ampia varietà di attività enzimatiche coinvolte nella loro sintesi, idrolisi e, in generale, nella loro modifica. Gli enzimi responsabili della degradazione dei carboidrati sono le glicosil idrolasi (glicosidasi), una classe di enzimi ubiquitari che catalizzano l'idrolisi dei legami glicosidici. Questi enzimi, che sono classificati in base alla loro specificità per la configurazione del carbonio anomero del substrato ( $\alpha$  o  $\beta$ -glicosidasi), vengono catalogati in famiglie e sottofamiglie in base alla loro sequenza. La banca dati che contiene questa classificazione si chiama **CAZy**, acronimo di "**Carbohydrate Active enZyme**" ([www.cazy.org](http://www.cazy.org)) (Cantarel et al. 2009). Ad oggi (Ottobre 2010) CAZy comprende 118 famiglie a loro volta raggruppate in 14 clan con strutture 3D simili. E' interessante sottolineare che quasi 1000 geni codificanti per glicosil idrolasi, non sono stati assegnati ad alcuna famiglia.

Le glicosil idrolasi si definiscono *inverting* o *retaining* in base al loro meccanismo catalitico (Koshland, 1953). Gli enzimi *inverting* idrolizzano i legami glicosidici con l'inversione della configurazione del carbonio anomero nel prodotto rispetto a quella del substrato. Al contrario, le glicosil idrolasi di tipo *retaining* (Figura 1) operano mediante un meccanismo catalitico a due stadi che comporta il mantenimento della configurazione del carbonio anomero del substrato. Entrambi i meccanismi, in genere, vedono protagonisti due residui carbossilici. In questo caso, durante il primo stadio di reazione (glicosilazione) il residuo carbossilico deprotonato dell'enzima funziona come nucleofilo ed attacca il carbonio anomero. Nello stesso tempo, l'allontanamento del gruppo uscente è assistito con una catalisi acida generale dal gruppo carbossilico protonato. Questo stadio termina con la formazione dell'intermedio glicosil-enzima. Nel secondo stadio (deglicosilazione), lo stesso gruppo carbossilico che prima ha funzionato da acido, ora deprotonato, attiva con una catalisi basica generale una molecola d'acqua, che può attaccare il carbonio anomero dell'intermedio, rilasciando così lo zucchero e l'enzima libero.



**Figura 1:** Meccanismo catalitico delle glicosil idrolasi *retaining*

Le glicosidasi possono essere impiegate utilmente in una varietà di applicazioni in campo industriale, per esempio nelle bioconversioni di prodotti polisaccaridici di scarto per ottenere prodotti di alto valore aggiunto, o, in campo molecolare e bio-medico, per lo sviluppo di nuovi farmaci a base carboidratica.

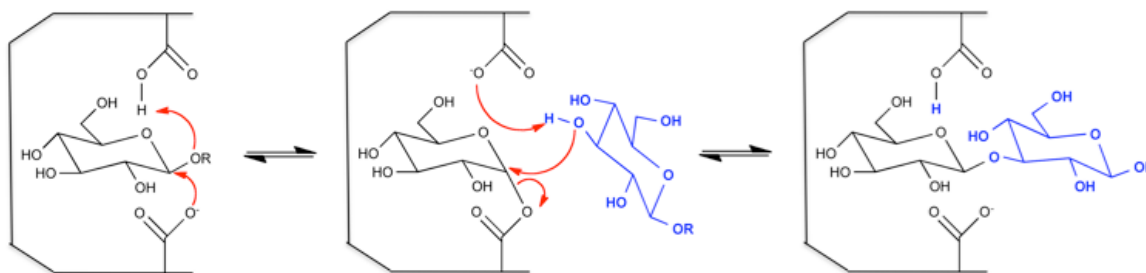
Tra le applicazioni industriali più importanti delle glicosidasi la conversione dei polisaccaridi di origine vegetale per ottenere bio-etanolo e precursori organici per sintesi chimica è tra le più note e studiate. La cellulosa e lo xilano, rispettivamente la più abbondante fonte di energia rinnovabile del pianeta e il principale costituente delle emicellulose, rappresentano, infatti, degli scarti virtualmente inesauribili per produrre zuccheri fermentabili. La completa depolimerizzazione di cellulosa e xilano a glucosio e xilosio richiede l'azione sinergica di diversi enzimi come cellulasi, xilanasi, cellobioidrolasi,  $\beta$ -glucosidasi e  $\beta$ -xilosidasi. Inoltre, glicosil idrolasi da microrganismi termofili e ipertermofili, essendo resistenti alle alte temperature, possono sopperire alla limitata stabilità degli enzimi tradizionali alle condizioni operative estreme di questi processi.

Molte glicosidasi, come le  $\alpha$ -D-mannosidasi, hanno anche importanti applicazioni in campo molecolare e biomedico, in quanto enzimi responsabili della maturazione degli oligosaccaridi mannosilati che, nelle glicoproteine, sono protagonisti della comunicazione cellulare (Davies, 2000). In gravi condizioni patologiche, come nelle metastasi tumorali, i livelli di mannosilazione sono alterati e gli inibitori delle  $\alpha$ -D-mannosidasi sono riconosciuti come utili farmaci (Goss et al. 1997). Inoltre, mutazioni nel gene dell' $\alpha$ -D-mannosidasi lisosomiale causano l' $\alpha$ -mannosidosi, una grave malattia genetica che genera progressivo ritardo mentale nei bambini affetti (Pittis et al. 2007).

Le glicosidasi si sono rivelate utili anche nella sintesi di oligosaccaridi. In campo farmaceutico, oligosaccaridi contenenti sequenze terminali  $\alpha$ -Gal-(1-3)- $\alpha$ -Gal (epitopo  $\alpha$ -gal) sono stati identificati come i maggiori antigeni responsabili del rigetto iperacuto (Galili 2001). La reazione immunologica mediata da anticorpi anti-Gal che legano specificamente l'epitopo  $\alpha$ -gal rappresenta la principale barriera negli xenotrapianti in uomo di organi da fonti non appartenenti all'ordine dei primati. Tra le strategie proposte per risolvere questo problema l'utilizzo di oligosaccaridi sintetici capaci di inibire gli anticorpi anti-Gal è tra le più promettenti (Macher and Galili 2008).

In campo alimentare i galattooligosaccaridi (GOS), costituiti principalmente da D-Gal- $\beta$ (1 $\rightarrow$ 3)-D-Gal (3-galattobiosio), D-Gal- $\beta$ (1 $\rightarrow$ 6)-Lac (6'-galattosil-lattose), D-Gal- $\beta$ (1 $\rightarrow$ 3)-D-Glc (3-galattosil-glucosio), e D-Gal- $\beta$ (1 $\rightarrow$ 3)-Lac (3'-galattosil-lattosio) (Martinez-Villaluenga et al. 2008), sono impiegati come prebiotici capaci di stimolare la proliferazione della microflora intestinale e sono presenti in diversi prodotti alimentari per l'infanzia (Coulier et al. 2009).

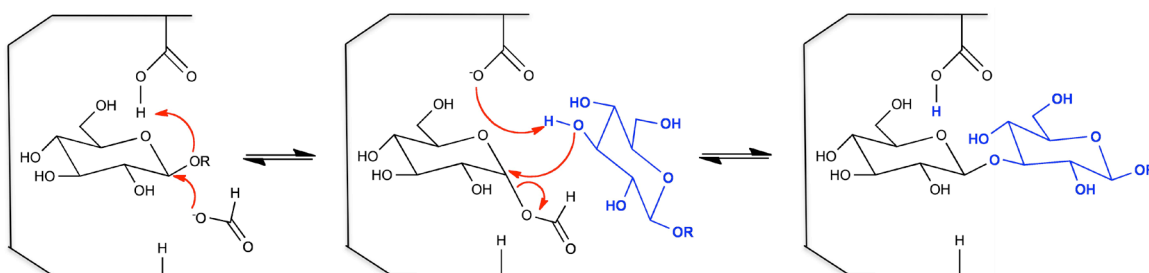
Tali approcci necessitano tuttavia della sintesi in grandi quantità di specifici oligosaccaridi rendendo indispensabile lo sviluppo e la messa a punto di nuove tecniche per la sintesi di oligosaccaridi e glicoconiugati. In un approccio classico, le glicosidasi retaining vengono utilizzate a questo scopo mediante reazioni di transglicosilazione. In questo caso, si utilizzano elevate concentrazioni di alcol o di un altro zucchero che permettono la risoluzione dell'intermedio covalente e la sintesi del legame glicosidico (Figura 2). In questa reazione i due substrati della reazione sono definiti *donor* e *acceptor*.



**Figura 2:** Meccanismo di transglicosilazione di una glicosil idrolasi *retaining*

Occorre tuttavia far presente che, nelle reazioni di transglicosilazione, il prodotto ottenuto, avendo una configurazione anomERICA uguale a quella del substrato, è a sua volta suscettibile all'idrolisi da parte dell'enzima stesso. Di conseguenza le rese finali non sono mai superiori al 40%.

Questo limite è stato superato per la prima volta, nel 1998, con l'avvento delle *glicosintasi* una nuova classe di enzimi ingegnerizzati (Mackenzie et al, 1998; Moracci et al, 1998; Malet e Planas 1998). Mutando il residuo nucleofilo di una  $\beta$ -glicosidasi retaining con un residuo non-nucleofilo (Ala, Gly, Ser), si produceva la completa inattivazione dell'enzima. Riattivando però il mutante con substrati opportunamente attivati si osservava la sintesi di oligosaccaridi. Tra le altre, suscitano notevole interesse le *glicosintasi* ottenute a partire da glicosil idrolasi termofile (Figura 3), che consentono di combinare le proprietà utili alla sintesi con la stabilità ad elevate temperature propria delle proteine di origine termofila.



**Figura 3:** Meccanismo di azione di una glicosintasi ottenuta da glicosil idrolasi termofila

Nonostante le glicosintasi offrano una valida e conveniente alternativa alla sintesi chimica degli oligosaccaridi, glicosil idrolasi recalcitranti a diventare glicosintasi sono tutt'altro che rare e, ad oggi, è stato possibile ottenere glicosintasi solo da 11 famiglie di glicosil idrolasi. Tra queste, recentemente, sono state sviluppate per la prima volta due  $\alpha$ -fucosintasi, a partire da enzimi della famiglia GH29 con un nuovo approccio che utilizza come *donor* la  $\beta$ -fucosil-azide (Cobucci-Ponzano et al. 2009). Nonostante gli sforzi effettuati nello studio nel campo della sintesi chemo-enzimatica dei carboidrati, l'intera linea di ricerca soffre al momento l'assenza di una comune strategia per convertire glicosil idrolasi *retaining* in glicosintasi. Per questo motivo lo studio e la dettagliata caratterizzazione dei meccanismi catalitici delle glicosil idrolasi rappresentano quindi un prerequisito indispensabile per lo sviluppo di nuove attività glicosintasiche.

Il lavoro svolto durante il mio dottorato di ricerca è stato dedicato all'identificazione e caratterizzazione di glicosil idrolasi per applicazioni biotecnologiche ed è articolato in

due parti. Nella prima parte mi sono occupato dell'applicazione delle glicosidasi alla sintesi degli oligosaccaridi. In particolare, mi sono dedicato allo studio e della caratterizzazione del meccanismo catalitico della  $\beta$ -galattosidasi dall'eubatterio moderato termofilo *Alicyclobacillus acidocaldarius* (Aa $\beta$ gal) per lo sviluppo di una nuova  $\beta$ -galattosintasi. Inoltre, ho caratterizzato l'attività sintasica di una nuova  $\alpha$ -galattosintasi, ottenuta dalla  $\alpha$ -galattosidasi del batterio ipertermofilo *Thermotoga maritima* (TmGalA), allo scopo di validare l'approccio basato sull'impiego di  $\beta$ -glicosil azidi recentemente proposto (Cobucci-Ponzano et al, 2009).

La seconda parte è dedicata allo studio dell'attività idrolitica di due glicosil idrolasi dal crenarchaeota ipertermofilo *Sulfolobus solfataricus*: la  $\beta$ -glicosidasi SSO1353 e l' $\alpha$ -mannosidasi (Ss $\alpha$ -Man). La caratterizzazione di SSO1353 è stata rivolta allo studio del meccanismo catalitico e in particolare all'identificazione dei residui nucleofilo e acido/base. Lo studio dell'attività di Ss $\alpha$ -Man è stato invece rivolto alla caratterizzazione della specificità di substrato di quest'enzima con particolare riguardo all'attività di de-mannosilazione di glicoconiugati e glicoproteine.

### **1.1: Caratterizzazione della $\beta$ -galattosidasi da *Alicyclobacillus acidocaldarius* (Aa $\beta$ gal)**

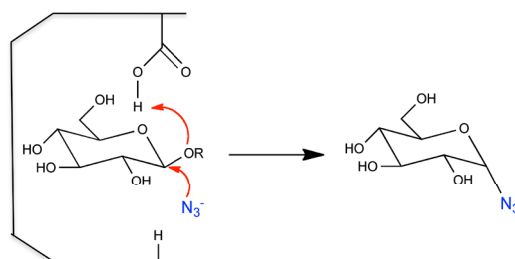
Aa $\beta$ gal è un enzima termofilo (temperatura di saggio 65°C) appartenente alla famiglia GH42 delle glicosil idrolasi. In questa famiglia è nota una sola struttura cristallografica, ottenuta della  $\beta$ -galattosidasi da *Thermus thermophilus*, caratterizzata da un folding di tipo *Tim Barrel* ( $\beta/\alpha$ )<sub>8</sub>, che catalizza l'idrolisi di legami  $\beta$ -galattosidici con un meccanismo di tipo *retaining*. I residui corrispondenti all'acido/base e al nucleofilo di Aa $\beta$ gal, rispettivamente Glu157 e Glu313, sono stati identificati recentemente (Di Lauro et al, 2008) ma il mutante nel nucleofilo non si comporta da glicosintasi. L'obiettivo della mia tesi è stato quello di caratterizzare nuovi mutanti di Aa $\beta$ gal per comprendere i motivi molecolari all'interno del sito attivo che potrebbero essere responsabili incapacità di sintesi e di sviluppare un diverso approccio per ottenerne una nuova  $\beta$ -galattosintasi.

#### **Risultati conseguiti**

Vista la mancanza di attività sintasica del mutante al residuo nucleofilo E313G, ho cercato, all'interno del sito attivo di Aa $\beta$ gal, la presenza di altri residui di natura nucleofila che potrebbero essere coinvolti nel meccanismo catalitico. Attraverso allineamento di sequenze multiple, con altre  $\beta$ -galattosidasi appartenenti alla famiglia GH42, ho identificato due residui carbossilici altamente conservati: Glu361 e Asp276. In seguito, mediante predizione di struttura tridimensionale, ottenuta tramite il programma 3D-JIGSAW, e successiva sovrapposizione con l'unica struttura tridimensionale disponibile per la famiglia GH42, la  $\beta$ -galattosidasi di *T. thermophilus*, ho potuto stabilire che questi due residui sono posizionati nel sito attivo dell'enzima. Per stabilire il ruolo dei residui Glu361 e Asp276, ho utilizzato il classico approccio che comprende la sostituzione, tramite mutagenesi sito-diretta, del residuo d'interesse, la successiva espressione e la caratterizzazione del mutante ottenuto.



Sono stati quindi preparati e purificati i mutanti E361G e D276G partendo dal gene del *wild type* clonato in pGEX2TK come proteina di fusione alla glutatione S-transferasi (GST) utilizzata per la purificazione mediante resina di affinità. Ottenuti i mutanti, ho saggiato la loro attività e misurato le costanti cinetiche per poi paragonarle al *wild type*. È stata così osservata una diminuzione del valore di  $k_{cat}$  di 48 e di 443 volte rispettivamente per i mutanti D276G ( $58.85\text{ s}^{-1}$ ) e E361G ( $6\text{ s}^{-1}$ ) se paragonata a quella dell'enzima *wild type* ( $2657\text{ s}^{-1}$ ). Per valutare se i residui Glu361 e Asp276 potessero avere o meno un ruolo diretto nel meccanismo catalitico in associazione con il residuo nucleofilo Glu313 e il residuo acido/base Glu157, ho condotto esperimenti di *chemical rescue*. Questa tecnica consente il recupero di attività enzimatica di un mutante in residui del sito attivo in presenza di un substrato dotato di un buon gruppo uscente e di una piccola molecola nucleofila esterna, come ad esempio sodio formiato o sodio azide. L'analisi del legame formatosi tra il nucleofilo e la parte gliconica consente di determinare la natura del residuo mutato. Se, infatti, il prodotto mostrerà, al carbonio anomero, una configurazione invertita rispetto al substrato di partenza, si avrà la prova che il residuo mutato è il nucleofilo della reazione (Figura 4).



**Figura 4:** *Chemical rescue* del mutante al residuo nucleofilo di una glicosil idrolasi *retaining* in presenza di azide.

Il mutante D276G non è stato riattivato in nessuna delle condizioni utilizzate, e ciò ha permesso di escludere un suo ruolo diretto nella catalisi. Al contrario, il mutante E361G ha mostrato una riattivazione di 18 e di 10 volte, in presenza, rispettivamente, di sodio formiato e sodio azide.

Per stabilire in che modo la presenza del nucleofilo esterno agisse nella catalisi e, conseguentemente, il ruolo del residuo Glu361 ho analizzato i prodotti di reazione del mutante E361G, in presenza di 2-NP- $\beta$ -D-galattopiranoside (2NP-Gal), nelle condizioni standard di reazione, sia in presenza di sodio azide che di sodio formiato. L'analisi, tramite cromatografia su strato sottile (TLC) dei prodotti di reazione del mutante E361G in presenza di sodio azide, ha mostrato, già nei primi minuti di reazione, la formazione di soli prodotti UV visibili che si accumulano dopo oltre 16 ore di incubazione. Tali prodotti sono il risultato del trasferimento di galattosio ad una molecola di 2-NP-Gal che funziona da *acceptor* (autocondensazione). Non è stato possibile rilevare, invece, la presenza della galattosil-azide attesa per *chemical rescue*. I prodotti osservati per TLC hanno un aspetto simile ai prodotti di transgalattosilazione ottenuti dall'enzima *wild type*. Tuttavia è da sottolineare però che Aa $\beta$ gal *wild type*, pur svolgendo reazioni di transgalattosilazione, per la sua elevata attività idrolitica ( $k_{cat}$ , su 2NP-Gal pari a  $2657\text{ s}^{-1}$ ), li idrolizza già nei primi 20 minuti di reazione impedendone l'accumulo. Al contrario, il mutante E361G è capace di transgalattosilare ma ha perso gran parte della sua attività d'idrolisi mostrando una  $k_{cat}$  di  $684\text{ s}^{-1}$  e di  $167\text{ s}^{-1}$  (in presenza di sodio formiato e sodio azide, rispettivamente) permettendo così l'accumulo di prodotti dopo 3 ore di reazione.

Questo risultato mi ha incoraggiato a saggiare, in condizioni standard, l'attività transgalattosidasi del mutante E361G in presenza di 50 mM NaN<sub>3</sub> con differenti *acceptor* ed in diversi rapporti (2:1, 4:1 e 8:1 ) utilizzando il 2NP-Gal come *donor*. Gli *acceptor* utilizzati sono stati il CH<sub>3</sub>-β-D-xilopiranoside, il 4NP-α-D-xilopiranoside e il 4NP-β-D-glucopiranoside. Ho quindi analizzato le reazioni per TLC che ha mostrato la sintesi di prodotti di transgalattosilazione in tutte le condizioni ed in particolare con gli *acceptors* CH<sub>3</sub>-β-D-xilopiranoside e 4NP-β-D-glucopiranoside. La TLC ha permesso, inoltre, di stabilire che il rapporto 2:1 è quello più favorevole a limitare il problema dell'autocondensazione. Per meglio confrontare l'attività transgalattosidica del mutante E361G e del wild type, sono stati condotti dei saggi sulle differenti coppie donor:acceptor precedentemente indicate utilizzando le stesse unità enzimatiche (1.62 U) per entrambi gli enzimi. L'analisi delle miscele di reazione mediante TLC ha confermato chiaramente, che il mutante E361G svolge la transgalattosilazione più efficientemente del wild type.

Sebbene il mutante del residuo nucleofilo E313G non mostri riattivazione in presenza di nucleofili esterni (Di Lauro et al, 2008), sorprendentemente, è stata osservata la sintesi di prodotti ad elevate concentrazioni di substrato (2NP-Gal 80 mM). La caratterizzazione mediante 1 H NMR, eseguita in collaborazione con la Prof. Corsaro dell'università di Napoli "Federico II", ha permesso di identificare i disaccaridi 2NP-β-gal-β(1,3)gal e 2NP-β-gal-β(1,4)gal. Questo risultato suggerisce per questo mutante un meccanismo di catalisi diverso dal wild type, nel quale, in seguito alla mutazione di E313 in glicina, un secondo residuo funziona da nucleofilo alternativo permettendo la formazione dell'intermedio covalente glicosil enzima. Per identificare questo residuo si è ricorso al metodo basato sull'impiego di un inibitore meccanismo-specifico, 2,4DNP-2-deossi-2-Fluoro-galattopiranoside (2,4DNP- 2F-Gal) e analisi di spettrometria di massa tramite nano-ESI-MS/MS. Questo tipo di approccio, applicato al mutante E313G, sfortunatamente non ha consentito di identificare il residuo alternativo, tuttavia in particolari condizioni di inibizione (rapporto molare Enzima:Inibitore = 1:10<sup>5</sup>) è stato possibile osservare, mediante spettrometria di massa nano-ESI-MS/MS, un forte segnale monocarica m/z 535 presente durante tutta l'analisi e compatibile con il prodotto di transglicosilazione dell'inibitore 2,4DNP-2F-Gal-(2F-Gal) in forma sodiata. Presumibilmente, nel mutante E313G, il residuo Glu157 continua a funzionare da catalizzatore acido/base permettendo l'attivazione dell'acceptor. Per ovviare a questo problema sono stati preparati tre doppi mutanti, E157G/S/Q-E313G. Tutti e tre gli enzimi sono risultati completamente inattivi alle condizioni standard di reazione. Successivi esperimenti di *chemical rescue* hanno mostrato riattivazione in presenza di sodio azide solo del doppio mutante E157Q-E313G contestualmente alla sintesi del prodotto β-GalN<sub>3</sub> confermando la formazione di un intermedio di reazione α-glicosil-enzima. Questo risultato indica chiaramente che il doppio mutante E157Q-E313G è il candidato ideale per identificare il residuo nucleofilo alternativo con l'inibitore meccanismo specifico 2,4DNP- 2F-Gal. Questo studio richiede ulteriori analisi.

## **1.2: Caratterizzazione dell'attività α-galattosintasi del mutante di TmGalA D327G**

Una α-galattosidasi da *Thermotoga maritima* MSB8 (TmGalA), classificata in CAZy come glicoside idrolasi di famiglia GH36 di cui è noto il meccanismo di reazione *retaining*, la struttura 3D ed i residui catalitici (nucleofilo e acido-base corrispondenti rispettivamente al residuo D327 e D387) (Liebl et al,1998; Comfort et al, 2007) è

stata utilizzata per produrre una nuova  $\alpha$ -galattosintasi. Scopo del mio lavoro è stato quello di applicare al mutante D327G, cataliticamente inattivo, l'approccio basato sull'impiego di  $\beta$ -glicosil azidi e caratterizzarne l'attività sintasica.

## Risultati conseguiti

Miscele di reazione contenenti il mutante TmGalA D327G e incubate a 65°C in 50 mM tampone sodio acetato a pH 5.0 in presenza di  $\beta$ -GalN<sub>3</sub> come unico substrato (a concentrazione tra 1 e 14 mM) non hanno mostrato nessuno prodotto di auto condensazione. Al contrario, utilizzando come acceptor 4NP- $\alpha$ -D-Glc, 4NP- $\alpha$ -Man, 4NP- $\alpha$ -Xyl, and 4NP- $\beta$ -D-Xyl con un rapporto molare 1:1 con il donore ( $\beta$ -GalN<sub>3</sub>) è stato possibile osservare diversi prodotti di sintesi. Successivamente è stata misurata l'efficienza di trngalattosilazione analizzando aliquote di reazione mediante High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) ed è stata caratterizzata la struttura dei prodotti ottenuti in collaborazione con la Prof. Corsaro dell'Università di Napoli "Federico II".

In presenza di 4NP- $\alpha$ -Glc è stata osservata la sintesi di un solo composto,  $\alpha$ -Gal-(1-6)- $\alpha$ -Glc-4NP, con una resa del 33%. Rese maggiori sono state ottenute con 4NP- $\alpha$ -Xyl and 4NP- $\beta$ -Xyl (40% e 38% rispettivamente) con la formazione di due prodotti specifici nel primo caso,  $\alpha$ -Gal-(1-2)- $\alpha$ -Xyl-4NP e  $\alpha$ -Gal-(1-4)- $\alpha$ -Xyl-4NP, ed un singolo prodotto,  $\alpha$ -Gal-(1-4)- $\beta$ -Xyl-4NP nel secondo caso. Nella reazione con 4NP- $\alpha$ -Man come acceptor sono stati isolati due prodotti corrispondenti all' $\alpha$ -Gal-(1-6)- $\alpha$ -Man-4NP e all' $\alpha$ -Gal-(1-3)- $\alpha$ -Man-4NP, con una resa totale del 51%.

Questi risultati confermano che l'approccio recentemente proposto per le glicosintasi ottenute da famiglia GH29 e basato sull'impiego di  $\beta$ -glicosil azidi da usare come substrato per la produzione di nuove  $\alpha$ -glicosintasi (Cobucci-Ponzano et al, 2009) è valido ed è estendibile anche a  $\alpha$ -glicosintasi derivate da altre famiglie di glicosil idrolasi (Cobucci-Ponzano et al. *in press*).

### 2.1: Caratterizzazione del residuo nucleofilo della $\beta$ -glicosidasi SSO1353 da *Sulfolobus solfataricus*

Recentemente, nel genoma di *S. solfataricus* (<http://www-archbac.u-psud.fr/projects/sulfolobus/>), è stata clonata, espressa e caratterizzata la ORF SSO1354 (Maurelli et al. 2008). L'enzima codificato è risultato una endoglucanasi attiva su xilano e cellulosa che può essere utilizzata nella degradazione di questi polisaccaridi per produrre zuccheri fermentabili. La ORF SSO1353, adiacente alla SSO1354 nel genoma di *S. solfataricus*, codifica per una proteina di funzione ignota. Nel laboratorio presso cui ho svolto la mia attività di ricerca, SSO1353 è stata espressa e caratterizzata biochimicamente dimostrandosi una  $\beta$ -glicosidasi specifica per gluco- e xilosidi (Cobucci-Ponzano et al, 2010). Questa glicosidasi è un enzima completamente nuovo non classificato in alcuna famiglia di CAZy. Il mio lavoro, relativo a questo nuovo enzima, è stato volto all'identificazione dei residui catalitici che funzionano da nucleofilo e da acido/base, mediante rispettivamente il metodo basato sull'impiego di un inibitore meccanismo-specifico e del *chemical rescue*.

## Risultati conseguiti

L'analisi mediante allineamento multiplo della sequenza di SSO1353 e di altre sequenze con identità >22% ottenute mediante BLAST, ha permesso di identificare quattro residui altamente conservati, E335, D406, D462 e D458 di cui sono stati prodotti i rispettivi mutanti in glicina (Cobucci-Ponzano et al, 2010).

Ho effettuato prove di inibizione della SSO1353 *wild type*, opportunamente purificata, incubandola con l'inibitore meccanismo-specifico 2,4-dinitrofenil-2-deossi-2-fluoro- $\beta$ -glucopiranoside (2,4DNP-2F-Glc), con un rapporto enzima:inibitore 1:1000 (mol/mol) ed ho saggiato l'attività dell'enzima, in condizioni standard, ad intervalli di 30 minuti per 4 ore di incubazione. La misura dell'attività specifica della SSO1353 ha mostrato un massimo livello di inibizione dopo 2 ore di trattamento con l'inibitore.

Un'aliquota di SSO1353 è stata, quindi, incubata per 2 ore in presenza di 2,4DNP-2F-Glc la miscela è stata successivamente acidificata (pH 2.0 in acido formico 5%) e trattata con pepsina in un rapporto pepsina:SSO1353 1:20 (w/w) per 30 minuti a 37°C, condizioni messe a punto precedentemente sullo stesso enzima in assenza dell'inibitore.

Le miscele costituite dalle proteine digerite (SSO1353 inibita e il controllo non inibito) sono state successivamente separate mediante HPLC, utilizzando una colonna C18, e analizzate per MS/MS. Conoscendo il peso molecolare della componente gliconica dell'inibitore (2FGlc) che forma l'intermedio covalente con l'enzima (164 Da), è stato possibile creare *ad hoc* un *database* interno che permettesse di identificare, mediante l'impiego della funzione *MS/MS ion search* del programma *MASCOT*, il residuo marcato dall'inibitore. Le analisi per spettrometria di massa sono state svolte in collaborazione con la Dott.ssa Pocsfalvi dell'Istituto di Biochimica delle Proteine del CNR.

L'analisi per spettrometria di massa della SSO1353 non inibita ha permesso l'identificazione di un peptide di 1779.85 Da corrispondente agli amminoacidi 332-348. Il medesimo peptide, presente nel digerito della SSO1353 inibita, ha mostrato, invece, una massa di 1943.89 Da. La differenza tra le masse di questi due peptidi è di 164.04 Da, ovvero la massa attesa per il 2FGlc legato.

La successiva frammentazione del peptide di 1943.89 Da, presente nello spettro della SSO1353 trattata con l'inibitore, ha permesso, grazie al database utilizzato per l'analisi, l'identificazione di Glu335 come residuo responsabile del legame all'inibitore. Ciò mi ha permesso di identificare, inequivocabilmente, residuo Glu335 come nucleofilo della reazione di questa nuova  $\beta$ -glicosidasi. (Cobucci-Ponzano et al, 2010).

Inoltre, la caratterizzazione, svolta in collaborazione con la Prof. Corsaro dell'Università di Napoli "Federico II", dei prodotti ottenuti mediante *chemical rescue* del mutante D462 in presenza di sodio azide come nucleofilo esterno, ha consentito di associare a tale residuo la funzione di acido/base.

L'identificazione dei due siti catalitici di SSO1353 ha contribuito sensibilmente alla creazione di una nuova famiglia di glicosil idrolasi, GH116 in CAZy consentendo di estendere le conoscenze acquisite relative a quest'enzima anche ad altri membri

della famiglia (Cobucci-Ponzano et al, 2010). Si sta valutando il coinvolgimento di SSO1353 e della endoglucanasi SSO1354 nella degradazione dei polisaccaridi.

## **2.2: Caratterizzazione dell'attività $\alpha$ -mannosidasi di Ss $\alpha$ -Man su oligosaccaridi *N-Linked***

Le  $\alpha$ -D-mannosidasi sono enzimi chiave coinvolti nel processamento dell'antenna glucidica mannosilata delle glicoproteine eucariotiche e sono classificate nelle famiglie GH38 e GH47. Gli enzimi di queste famiglie differiscono nella struttura tridimensionale, nel meccanismo di reazione e nella specificità degli inibitori. Gli enzimi della famiglia GH38 sono inibiti dalla swainsonina e seguono un meccanismo *retaining* mentre quelli della famiglia GH47 sono inibiti dalla 1-deoxymannojirimicina e seguono un meccanismo tipo *inverting*. La famiglia GH38 comprende  $\alpha$ -D-mannosidasi di classe II che, in eucarioti, si trovano nel Golgi, nel citosol e nel lisosoma (Kawar et al. 2001). Lo studio di nuove  $\alpha$ -D-mannosidasi con particolare specificità per oligosaccaridi mannosilati e per l'antenna glucidica di glicoproteine può essere estremamente utile per progettare nuovi approcci molecolari per la modifica di questi importanti regolatori della comunicazione cellulare. Il gene di *Sulfolobus solfataricus* SSO3006, codificante una  $\alpha$ -D-mannosidasi (Ss $\alpha$ -Man), è stato clonato in *E. coli* nel laboratorio dove svolgo la mia attività di dottorato e la proteina ricombinante codificata (rSs $\alpha$ -Man) è stata espressa e purificata così come quella nativa (nSs $\alpha$ -Man). Entrambe le varianti sono state caratterizzate biochimicamente ed enzimaticamente. Il mio obiettivo è stato quello di caratterizzarne l'attività su oligosaccaridi del tipo *N-Linked*, in particolare Man<sub>7</sub>GlcNAc<sub>2</sub> e Man<sub>3</sub>GlcNAc<sub>2</sub>, e verso una glicoproteina standard quale la Ribonucleasi B.

### **Risultati conseguiti**

Ho saggiato l'attività di nSs $\alpha$ -Man (10,64  $\mu$ g, 96  $\mu$ mol) su oligosaccaridi *high-mannose* Man<sub>3</sub>GlcNAc<sub>2</sub> e Man<sub>7</sub>GlcNAc<sub>2</sub> (1200 pmol) in condizioni precedentemente ottimizzate (tampone sodio acetato 50 mM pH 5.5, 1mM ZnCl<sub>2</sub> incubati a 50°C) e prelevando aliquote ad intervalli di tempo regolari, rispettivamente 0, 1h, 4h, 16h. I prodotti di reazione sono stati analizzati mediante cromatografia di tipo HPAE-PAD con una colonna Carbopac PA200 applicando un gradiente di sodio acetato che ho appositamente ottimizzato per questa separazione.

L'analisi dei prodotti di reazione ha permesso di osservare l'attività  $\alpha$ -mannosidasi verso entrambi i substrati saggiati. L'analisi delle miscele di reazione, al variare del tempo, ha consentito di apprezzare, a fronte della totale scomparsa del substrato iniziale dopo 16 h di reazione, la progressiva comparsa di segnali compatibili con le varie glicoforme intermedie fino al chitobiosio (GlcNAc  $\beta$ -(1,4)-GlcNAc). L'attività dell'enzima su oligosaccaridi *high-mannose* mi ha suggerito di saggiare Ss $\alpha$ -Man verso una glicoproteina modello quale la Ribonucleasi B.

RNasi B, purificata da pancreas bovino, è una glicoproteina di circa 15 kDa con un singolo sito di glicosilazione (Asn34) ma si presenta in cinque diverse glicoforme, da Man<sub>5</sub> a Man<sub>9</sub>. Ho trattato un quantitativo pari a 1200 pmol di RNasi B con 95 pmol di nSs $\alpha$ -Man nelle condizioni precedentemente stabilite. L'analisi della miscela di reazione mediante cromatografia HPAE-PAD dopo 16 ore di incubazione, mi ha permesso di rilevare un netto aumento della quantità di mannosio libero suggerendo

così un'attività mannosidasica ad opera di nSs $\alpha$ -Man verso l'RNasi B. Quest'ipotesi è stata confermata analizzando, mediante spettrometria di massa SELDI-TOF, l'RNasi B trattata con nSs $\alpha$ -Man e quella non trattata. Lo spettro di massa del campione trattato con  $\alpha$ -mannosidasi ha infatti evidenziato la presenza di un nuovo segnale, m/z 14731.5, compatibile con l'RNasi B avente la glicoforma Man<sub>4</sub>GlcNAc<sub>2</sub> ottenuta presumibilmente dalla demannosilazione della variante Man<sub>5</sub>GlcNAc<sub>2</sub> in quanto tutte le altre glicoforme restano apparentemente inalterate. Non è al momento chiaro se questo dato indichi un'alta specificità di Ss $\alpha$ -Man per una singola forma glicosilata di RNasi B o derivi dalle specifiche condizioni di reazione applicate per quest'esperimento. Altri esperimenti saranno necessari per la caratterizzazione di quest'attività. I dati finora ottenuti, tuttavia, suggeriscono che Ss $\alpha$ -Man potrebbe essere coinvolta nell'idrolisi e nella maturazione delle glicoproteine *in vivo*.

## Abbreviations

2- or 4-NP	2- or 4-nitrophenyl
2,4DNP-2F-Gal	2,4-dinitrophenyl- $\beta$ -D-2-deoxy-2-fluoro-galactopyranoside
2,4DNP-2F-Glc	2,4-dinitrophenyl- $\beta$ -D-2-deoxy-2-fluoro-glucopyranoside
Aa $\beta$ -gal	$\beta$ -galactosidase from <i>Alicyclobacillus acidocaldarius</i>
CAZy	Carbohydrate Active enZYme classification
CH <sub>3</sub> -Xyl	Methyl- $\beta$ -D-xylopyranoside
DMSO	Dimethyl sulfoxide
DNJ	1-deoxymannojirimycin
Gal	galactopyranoside
GH	glycoside hydrolase
Glc	glucopyranoside
GOS	Galactooligosaccharides
GS	glycosynthase
GST	Glutathione S-Transferase
GT	glycosyltransferases
HPAEC–PAD	High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection
Man	mannopyranoside
nano-ESI-MS/MS	nano-electrospray ionization tandem mass spectrometry
ORF	open reading frame
SELDI-TOF	Surface-enhanced laser desorption/ionization-Time of flight
Ss $\alpha$ -man	$\alpha$ -mannosidase from <i>Sulfolobus solfataricus</i>
TLC	thin layer chromatography
TmGalA	$\alpha$ -galactosidase from <i>Thermotoga maritima</i>
Xyl	xylopyranoside
$\beta$ -GalN <sub>3</sub>	$\beta$ -galactosyl-azide

## **General introduction**



The synthesis and degradation of carbohydrates in the form of di-, oligo-, and polysaccharides, is crucial in all living domains. In fact they have vital importance as sources of energy, cellular structure and intracellular communication. Carbohydrates are the main source of energy in heterotrophic organisms, as polysaccharides, such as starch and glycogen, used as energy storage, which are subsequently hydrolysed to monosaccharides capable to enter in the metabolic cycle. Interestingly, carbohydrates in the form of glycoconjugates (glycoproteins and glycolipids) are responsible of important biological functions including cell-cell interactions, signal transduction, compartmentalization of proteins and antigenic response (Moremen, 2002; Staudacher et al. 1999).

### The Glycoside Hydrolases

Two main classes of catalysts are involved in the modification of carbohydrates in nature: Glycoside Hydrolases (GHs) and Glycosyltransferases (GTs), which are responsible for the hydrolysis and the synthesis of the sugars, respectively.

GTs catalyse the transfer of sugar moieties from activated *donor* molecules to specific *acceptor* molecules, forming glycosidic bonds. The IUB-MB (International Union of Biochemistry and Molecular Biology) classification of Glycosyltransferases is presently based on the specificity for the donor and acceptor substrates and do not indicate the intrinsic structural features of the enzymes. Moreover, Glycosyltransferases can be classified as either *retaining* or *inverting* according to the stereochemistry of the substrates and reaction products (Sinnott, 1990) and as members of the Leloir and the non-Leloir pathway on the basis of their substrate specificity. The formers use sugar nucleotides as *donors*, while the enzymes of the non-Leloir pathway exploit as *donors* glycosyl phosphates.

GHs (also named Glycosidases) are a widespread group of enzymes that hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. The IUB-MB Enzyme nomenclature of GHs is based on their substrate specificity and molecular mechanism. They are classified regarding the stereospecificity of the anomeric carbon in the substrate distinguishing them in  $\alpha$ - and  $\beta$ - glycosidases. Their activity over long glycosidic chains allows to distinguish between exo- or endo-glycosidases if they attack the glycosidic bonds at the ends of the substrate or inside the chain, respectively. Finally, the classification in *inverting* and *retaining* is based on the two catalytic mechanisms proposed by Koshland in 1953 (see below).

### The CAZy Database

Online since 1998, CAZy ([www.cazy.org](http://www.cazy.org)) is a database dedicated to the display and analysis of genomic, structural and biochemical information on **Carbohydrate-Active Enzymes** (CAZymes): Glycoside Hydrolases (GHs), GlycosylTransferases (GTs), Polysaccharides Lyases (PLs) and Carbohydrate Esterases (CEs). These enzymes are classified in families on the base of their aminoacidic sequence and some families can be further grouped in 'clans' showing conserved 3D-structures (Cantarel et al. 2009). As far as October 2010 in CAZy are allocated 118 families and 14 clans and it is continuously updated. In fact, the entries that are not classified are computed in family GH0. They include all the sequences that was not possible to

associate to existing families, or to reunite in a new GH family, because of lack of biochemical and enzymatic information.

The importance of this classification is that the catalytic machinery, the stereospecificity and the reaction mechanism (i.e. *inverting* or *retaining*) are conserved for all the GHs belonging to a certain family (Gebler et al. 1992), therefore, CAZy demonstrated to be extremely useful to predict these characteristics for any new glycosidase not yet characterized. It is clear that the characterization of a novel GH and, eventually, the creation of a new family, always represent a crucial increment in the understanding of this class of enzymes for both basic and applied research.

### **Biotechnological application of Glycoside Hydrolases**

The Glycoside Hydrolases play a key role in many biomedical applications, such as the production of universal blood by  $\alpha$ -N-acetylgalactosidase and  $\alpha$ -galactosidase hydrolyzing the A and B antigens respectively, to form the common H structure found in the O group (Liu et al. 2007); in industrial applications, such as baking and the brewing process, and in textile industry where they used different classes of GH for the treatment of fabrics in order to improve the softness, remove the pilling and creation of washout effects on blue denim. The main industrial applications of Glycoside Hydrolases, which may be considered as historical blockbusters, concern, however, the production of High Fructose Corn Syrup (HFCS), the formulation of laundry detergents, the production of bioethanol and synthesis of oligosaccharides.

The process by which HFCS is produced was first developed in 1957 (Marshall & Kooi, 1957). High-fructose corn syrup is produced by milling corn to produce corn starch (corn syrup). The corn syrup then is treated with  $\alpha$ -amylase to produce shorter chains of sugars called dextrins. This intermediate product is then treated with glucoamylases that convert dextrins to glucose in a process called saccharification. The result is glucose corn syrup, most of which (90%) is enzymatically converted to HFCS by glucose isomerase. Today the fructose syrup has replaced sucrose in many foods and mainly in almost all soft drinks. In 2008 the average consumption of HFCS for each person was approximately 17.1 kg in the United States (Economic Research Service U.S. Department of Agriculture). This indicates how many tonnes of  $\alpha$ -amylase annually are produced and used in bioreactors for the saccharification process.

If viewed on the basis of tonnage, the main application by far of enzymes is still in the laundry detergent sector, which includes about 30-40% of the total (Bommarius & Riebel, 2004).

In the laundry detergents, GH mainly used are the amylases and cellulases. Amylases facilitate the removal of starch-containing stains such as those from pasta, potato, gravy, chocolate and baby food, hydrolyzing the starch into dextrins and oligosaccharides. Cellulases are widely used in washing cotton fabrics to avoid the loss of colour and brightness of the tissues during use. They also prevent pilling, removing the small fibres without apparently damaging the major fibres and restores the fabric to its "as new" condition, and they aids the cleaning process decreasing the ability of cellulosic fibres to bind the soil.

Bioethanol, unlike petroleum, is a form of renewable energy that can be produced from agricultural feedstocks. It can be made from hydrolysis, by amylases, of carbohydrates source as sugar cane, potato, manioc, several cereals and food byproducts, as molasses, to obtain simple fermentable sugars (first generation bioethanol). Differently, cellulosic-ethanol (second generation bioethanol) is a type of bioethanol produced from lignocellulose (composed mainly of cellulose, hemicellulose and lignin), a structural material that comprises much of the mass of plants, by treatment with cellulase, xylanase and hemicellulase enzymes. The byproducts of lawn and tree maintenance are some of the more popular cellulosic materials for bioethanol production. This has the advantage of abundant and diverse cheap raw material compared to food source but requires a greater amount of processing to make the sugar monomers available to the microorganisms used to produce ethanol by alcoholic fermentation (Waclawovsky et al. 2010). These monosaccharides, and in particular the furan derivatives, also have the potential to serve as substitutes for the petroleum-based building blocks that are currently used in the production of plastics and fine chemicals (Román-Leshkov et al. 2006; Yang et al. 2010).

Complex carbohydrates are tools able to inhibiting specific enzymes responsible for viral infections. The action of a major viral envelope protein, neuraminidase, is to cleave the sialic acid from the membrane glycolipid so the new virus particles can be released from host cells. The specificity of this neuraminidase has been the target for the development of two successful influenza drug therapies derived from acid N-acetyl-D-neuramic (DANA) (Relenza™ GlaxoSmithKline, and Tamiflu™ Roche), that bind specifically to the neuraminidase active site, and thus inhibit the transmission of the virus (Packer et al. 2008).

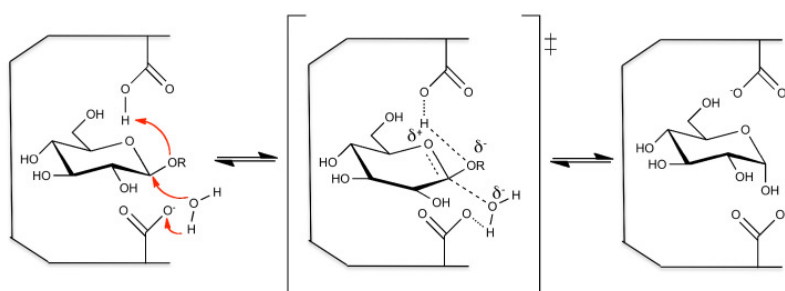
Heparin, the oldest carbohydrate-based drug and in the past isolated from animal organs, has been used clinically as an antithrombotic agent since the 1940s. This drug, in the past commercialized as a highly heterogeneous mixture of polysaccharides, then as chemically or enzymatically fragmented low-molecular-weight heparins, is associated with severe side effects, including heparin-induced thrombocytopenia, bleeding and allergic reactions. In the early 1980s the specific pentasaccharide responsible for the anticoagulant property was identified and a remarkable effort lasting more than 10 years was performed to understand the molecular bases of this property aiming to prepare synthetic oligosaccharides. As a result of this drug-development study, a synthetic pentasaccharide known as Arixtra™ (GlaxoSmithKline), which lowered the risks for heparin-induced thrombocytopenia has been available since 2002 (Seeberger & Werz, 2007).

The considerable role of oligosaccharides in several biological processes and their synthesis and subsequent applications as possible anti-cancer vaccines and antiviral drugs (Freire et al. 2006), represents a wide field, largely unexplored, with large interest in both basic and applied research. In these regards, GH as potential biotechnological tools are particularly appealing for their availability and wide specificity. Moreover, the interest in the identification and characterization of new glycosidases, is highly motivated by their biodiversity, which, as yet, has allowed to revealing novel activities for a wide range of substrates. In addition, the study of the structure/function relationship and reaction mechanism of GH is essential to

understand their role *in vivo* and to develop new methods for oligosaccharide synthesis and modification.

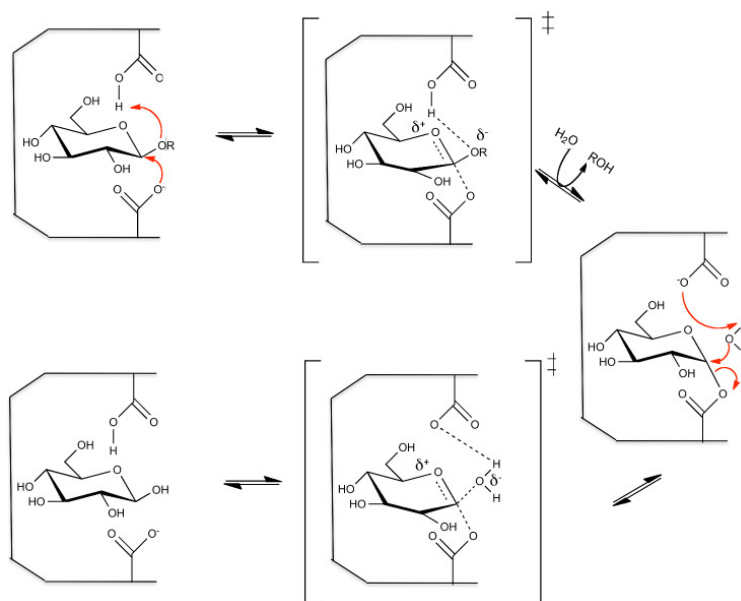
### Catalytic mechanisms of Glycoside Hydrolases

GHs follow two distinct mechanisms that are termed *inverting* or *retaining* if the enzymatic cleavage of the glycosidic bond liberates a sugar hemiacetal with the opposite or the same anomeric configuration of the glycosidic substrate, respectively. *Inverting* glycosidases (Figure 1.1) operate with a one step, single-displacement mechanism, involving oxocarbenium ion-like transition states, with the assistance of a general acid and a general base group, normally glutamic or aspartic acids that are typically located 6-11 Å apart, in the active site (McCarter & Withers, 1994).



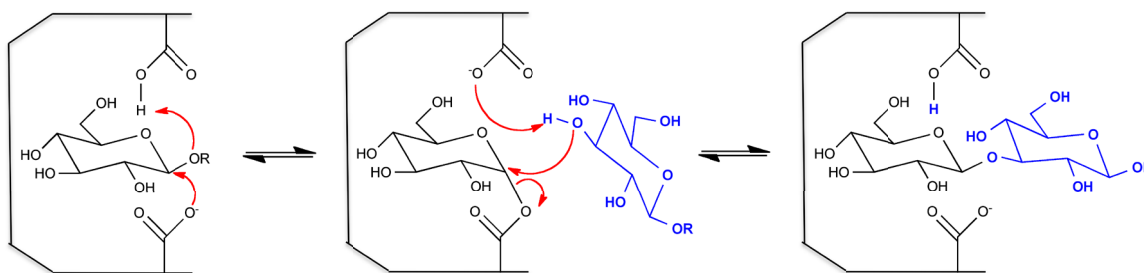
**Figure 1.1:** Catalytic mechanism of an *inverting*  $\beta$ -glycoside hydrolases

Instead, *retaining* enzymes (Figure 2.1) follow a two-step mechanism, again through an oxocarbenium ion-like transition state, but with formation of a covalent glycosyl-enzyme intermediate. The carboxyl group in the enzymatic active site functions as a general acid/base catalyst, and the carboxylate functions as the nucleophile of the reaction (Koshland, 1953). These residues are typically located 5.5 Å apart. In the first step, named glycosylation step, the nucleophile attacks the anomeric carbon of the substrate, while the other, acting in this step as a general acid, protonates the glycosidic oxygen, thereby assisting the leaving of the aglycon moiety. The concerted action of the two amino acids leads to the formation of a covalent glycosyl-enzyme intermediate (Sinnott, 1990) (McCarter & Withers, 1994). In the second step (known as the deglycosylation step), the glycosyl-enzyme intermediate is hydrolyzed, with the other residue now acting as a base catalyst deprotonating a water molecule that attacks the covalent intermediate. The product of the reaction retained the anomeric configuration of the substrate.



**Figure 2.1:** Catalytic mechanism of a *retaining*  $\beta$ -glycoside hydrolases

When an *acceptor* different from water, such as an alcohol or a sugar, intercepts the reactive glycosyl-enzyme intermediate, *retaining* enzymes work in transglycosylation mode (Figure 3.1). This property makes *retaining* GH interesting tools for the synthesis of carbohydrates as they usually allow the total control of the stereospecificity and, have good regiospecificity. However, typically, *retaining* glycosidases give only modest synthetic yields (10-40%) because the product maintains the same anomeric configuration of the substrate and can be hydrolyzed. This is the main obstacle for the exploitation of GH working in transglycosylation mode in large-scale synthesis of oligosaccharides (Perugino et al. 2005).



**Figure 3.1:** Transglycosylation reaction of a *retaining*  $\beta$ -glycoside hydrolases

It is worth noting that an increasing number of new glycosidases showed atypical mechanisms. Enzymes in families GH4 and GH109 operate through an NAD-dependent hydrolysis mechanism and are active both on  $\alpha$ - and  $\beta$ -glycosides (Liu et al. 2007; Yip et al. 2007). GH from family GH33 have a tyrosine acting as nucleophile instead of the canonical carboxylic acids (Watts et al. 2006), while, in family GH1, myrosinases have only the nucleophile and the deglycosylation step is provided by the co-enzyme L-ascorbate (Burmeister et al. 2000). Finally, in several GH families,

namely 18, 20, 23, 25, 56, 84, and 85, the nucleophile is missing and a 2-acetamido group of their substrate acts as an intramolecular nucleophile in a substrate assisted catalysis (Van Scheltinga et al. 1995). This indicates that much remains to be discovered in the continuously increasing GH families.

### Identification of the catalytic residues of glycoside hydrolases

With the increasing numbers of glycosidase sequences available from complete genomes, the characterization of these enzymes is essential to assign to them a function. In these regards, the identification of active site amino acid residues is fundamental and several approaches are available. First, it should be determined if the enzyme follows an *inverting* or *retaining* reaction mechanism. This can be easily elucidated by isolating the reaction products or by following the stereochemical course of the reaction by  $^1\text{H}$  NMR analysis (Withers et al. 1986). Then, the identification of the catalytic nucleophile and acid/base catalysts of *retaining* GH can follow different approaches.

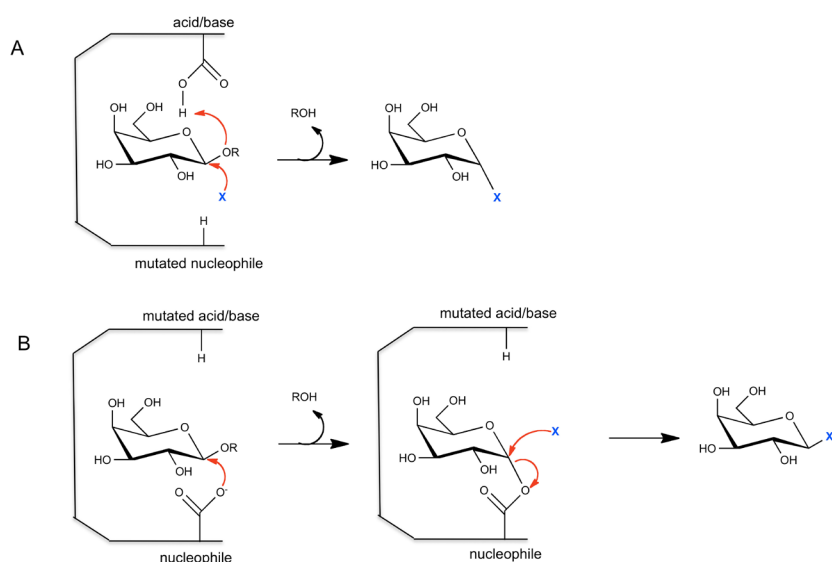
If 3D structural information at atomic resolution is available, the identities of active site amino acid residues can often be determined by detailed examination of the active site region. Conclusive identification of the active site residues frequently requires determination of the structure of an enzyme/inhibitor or enzyme/substrate complex. In these structures, the residues that are important for catalysis or, less directly, through binding of the substrate will be immediately apparent. However, 3D-structures are still not available for the vast majority of glycosidases and, even for those whose structures are known, obtaining structural information on enzyme/ligand complexes can be difficult. In fact, even when the identities of the residues in close spatial proximity to the substrate are known, their specific roles in catalysis frequently cannot be predicted. Thus, although these methods can suggest catalytic residues, in themselves they are not conclusive. Further insights into the roles in catalysis are usually obtained through site-directed mutagenesis, to replace conserved residues by nonionizable amino acids, followed by detailed kinetic analysis of the mutants produced.

The sequence of a GH of interest is aligned with other homologs and the resulting multi-alignment is analyzed for conserved charged residues (typically Asp, Glu, and His) or with polar side group (typically Ser and Tyr) that are predicted to be catalytic. This first screening allows to restrict the number of amino acids that can be modified in non-nucleophilic residues (typically Ala or Gly) by site-directed mutagenesis. The mutation of both the essential residues of retaining GH is expected to produce enzymes with a turnover number decreased by about  $10^4$ - $10^5$ -fold, if compared to the wild type. In particular, while the removal of the nucleophile would produce the complete loss of activity, because the mutant is unable to form the covalent glycosyl-enzyme intermediate (see Figure 2.1), the mutation of the acid/base would result of almost complete inactivation with substrates with poor leaving groups whereas the hydrolysis of substrates with good leaving group should be less affected. This is because substrates with good leaving groups need less acid assistance by the enzyme in the glycosylation step. Another consequence of the mutation of the acid/base will therefore be very low  $K_M$  values for such substrates, reflecting substantial accumulation of the glycosyl-enzyme intermediate.

Another diagnostic tool is the pH dependence of the mutant compared with that of the WT enzyme. The pH profiles of glycosidases are typically bell-shaped, mainly

reflecting the ionization state of the two carboxylic catalytic residues in the active site (Collins et al. 2005). Replacement of catalytic components usually alters the corresponding ionization in the pH profile. In particular, if the acid/base catalyst is removed, the basic (high-pH) limb of the profile should be severely affected and likely removed.

Numerous examples of this approach exist, but, frequently, only a incomplete evaluation of catalytic behaviour of mutant enzymes is performed; therefore, a third diagnostic tool and probably the most definitive is needed. This is the so-called *chemical rescue* of the activity for such mutants in the presence of suitably nucleophilic anions such as azide or formate. Briefly, the addition to the reaction mixture of small external anions, such as sodium azide, in the presence of an activated substrate, can lead to the reactivation of the enzymatic activity. If this is the case, the products of the reaction can be isolated and characterized to determine their stereochemistry, allowing the unequivocal identification of the function of the residue mutated (Zechel et al. 2001). In fact, in the case of the mutation of the nucleophile, the activity rescued with azide leads to glycosyl-azide products with the anomeric configuration *inverted* if compared to that of the substrate: i.e. production of  $\alpha$ -glucosyl-azide by mutated  $\beta$ -glucosidases (Figure 4.1A).

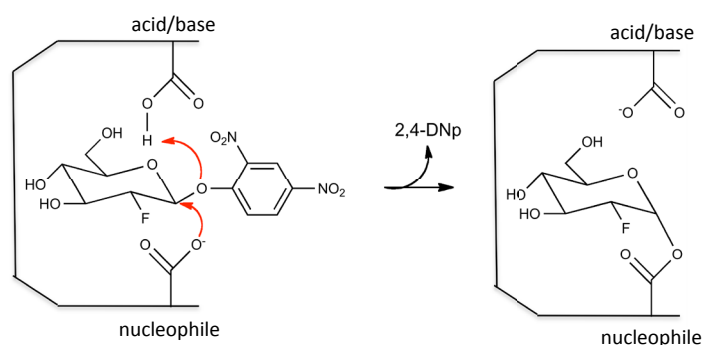


**Figure 4.1:** Mechanism of chemical rescue of the enzymatic activity of the mutants in the nucleophile (A) and the acid/base of the reaction (B). R: aglycon group, X: small external nucleophile.

Instead, the product with *retained* anomeric configuration (i.e.  $\beta$ -glucosyl-azide product from a mutant  $\beta$ -glucosidase) indicated that the general acid/base residue has been replaced (Figure 4.1B). The reason for this behaviour is that the azide ion replaces either the nucleophile residue in the first step of the reaction or the general base in the second step, respectively, depending on the mutant analyzed.

Another approach commonly exploited to identify the nucleophile in *retaining* GH is that combining a mechanism-based inhibitor and peptide mapping by mass

spectrometry. Briefly, mechanism-based inhibitors, as 2-deoxy-2-fluoroglycosides bearing good leaving groups, are ligands that bind to the active site by competing with the substrate (Withers & Aebersold, 1995). These molecules bind covalently to the enzyme and the fluoride at C2 serves to destabilize the oxocarbenium ion-like transition states, thereby slowing both the glycosylation and deglycosylation steps. The presence of a good leaving group (typically fluoride or dinitrophenolate) ensures that the intermediate is kinetically accessible. As a consequence, the intermediate accumulates (Figure 5.1). Successively, comparative peptide mapping of proteolytic digests allows the identification of the labelled peptide whose sequencing locates unequivocally the catalytic nucleophile.



**Figure 5.1:** Mechanism-based inhibitor reaction in retaining glycoside hydrolase active site.

The identification of the acid and the base catalysts in *inverting* GH also follows similar approaches including 3D-structure inspection, if available, alignments of the amino acid sequences of homologs to modify, by site-directed mutagenesis, potentially catalytic residues, and detailed kinetic characterization of the mutants (Hancock et al. 2006).

The determination of the catalytic residues of a GH is a fundamental step toward the exploitation of this enzyme in biotechnology. In fact, it is worth mentioning that the chemical rescue was propaedeutical to the development of novel methods for the chemo-enzymatic synthesis of oligosaccharides.

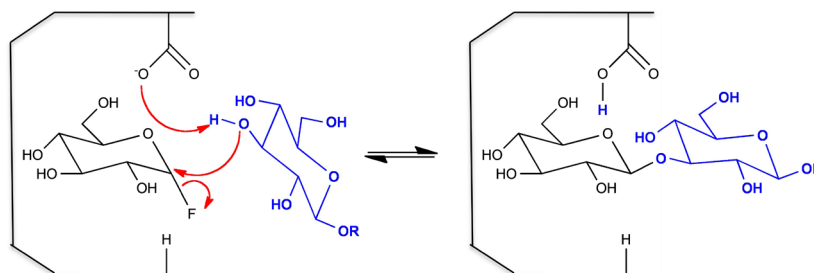
### The glycosynthases approach

To obtain an efficient method for the enzymatic oligosaccharides synthesis, several years ago was introduced a new class of mutant glycosidases, named glycosynthases (Mackenzie et al. 1998; Moracci et al. 1998; Malet & Planas, 1998). Glycosynthases derived from *retaining* glycosidases, in which, the active site nucleophile was replaced with a non-nucleophilic residue (as glycine or alanine) and in the presence of activated glycosides and suitable reaction conditions, the mutants synthesized oligosaccharides without hydrolysing them. As for glycoside hydrolases, also glycosynthase have been classified into *inverting* and *retaining* (Figure 6.1 and



7.1).

Inverting glycosynthase use a glycosyl fluoride with the opposite anomeric configuration to that of the normal substrate ( $\alpha$ -F-glucoside for mutant  $\beta$ -glucosidases). Reactivation then occurs via transglycosylation, resulting in oligosaccharides that cannot be hydrolysed by the glycosynthase and accumulate. In the first reported glycosynthase, Glu358 of the  $\beta$ -glucosidase from *Agrobacterium sp.* was replaced by an alanine residue (Figure 6.1) (Mackenzie et al. 1998).



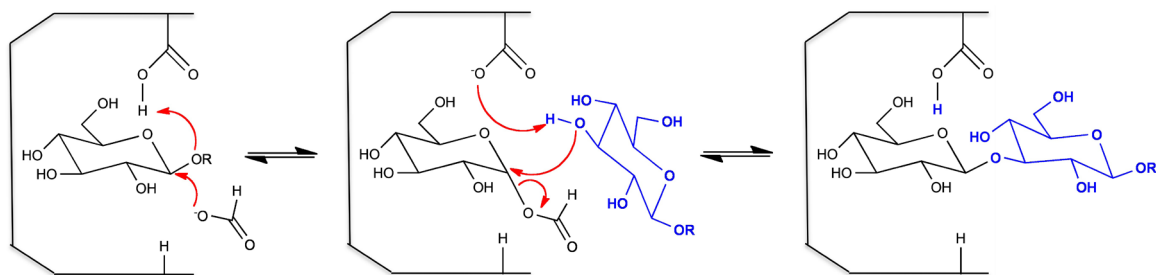
**Figure 6.1:** Reaction mechanism of *inverting* glycosynthase

In this case the product of the reaction with the  $\beta$ -anomeric configuration could not be hydrolysed by the mutant enzyme, which was inactivated by mutation. This reaction led to oligosaccharide products with yields typically >60% (Mackenzie et al. 1998). This invention enabled the production of  $\beta$ -1,4-linked cello-oligosaccharides that can be used as cellulase inhibitors and substrates. This method was successively improved by replacing the residue Glu358 with a serine or a glycine catalyzing the transfer of sugars with final yields > 63% and >83%, respectively (Mayer et al. 2000)(Tolborg et al. 2002).

Following a similar approach, Planas and co-workers, developed several endoglucanases acting as glucansynthases (endo-glycosynthases), by transferring oligosaccharides of varying degrees of polymerization to sugar *acceptors*. This was a valuable tool for the synthesis of complex oligosaccharides and polysaccharides showing highly specificity and efficiency with a final yield of 76% (Malet et al. 1998; Fairweather et al. 2002).

*Retaining*  $\beta$ -glycosynthases, instead, synthesize oligosaccharides by using activated *donors* with the same anomeric configuration of the substrate (typically 2-, 4-nitrophenyl- or 2,4-dinitrophenyl-glycosides) in the presence of external ions such as sodium formate (Figure 7.1) (Trincon et al. 2000).

-



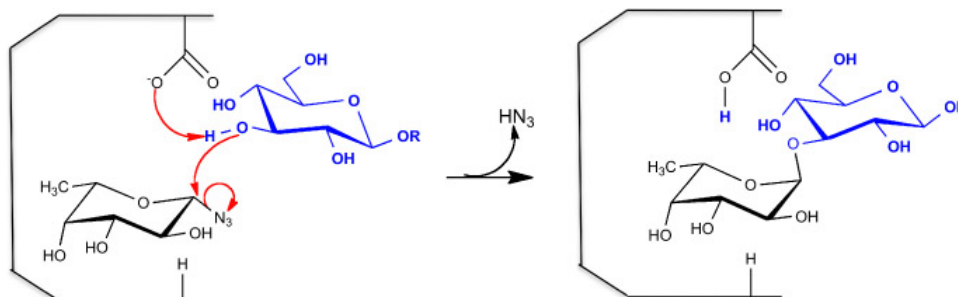
**Figure 7.1:** Reaction mechanism of *retaining* glycosynthase

This external ion mimics the natural active site carboxylate promoting the formation of a metastable glycosyl-formate intermediate. The products have the same anomeric configuration of the *donor*, but they cannot be hydrolysed by the enzyme, even in the presence of sodium formate, because of the bad leaving ability of the oligosaccharide groups (high  $pK_a$ ). Therefore they accumulated in the reaction mixtures leading to quantitative yields. However, it is worth noting that only mutant glycosidases from hyperthermophiles act efficiently as *retaining* glycosynthases in the presence of formate (Moracci et al. 1998). Presumably, the intermediate formyl-glycoside, which has been identified in one case (Viladot et al. 2001), does not react efficiently with sugar *acceptors* in mesophilic enzymes. The first example of *retaining*  $\beta$ -glycosynthase was obtained from  $\beta$ -glycosidase from *Sulfolobus solfataricus* (Ss $\beta$ -gly E387G) (Moracci et al. 1998). The Ss $\beta$ -gly mutant E387G synthesizes  $\beta$ -1,3- or  $\beta$ -1,6-linked tetrasaccharides, the building blocks of the  $\beta$ -1,3-1,6-glucans that are recognized as elicitors of the defence response against pathogens in plants and invertebrates (Perugino et al. 2004) (Côté et al. 1994).

The approach leading to *inverting* glycosynthases resulted suitable for both exo- and endo-glycosidases and has successfully been applied to a variety of GHs belonging to several families of the CAZy classification, namely, GH1, GH2, GH5, GH7, GH10, GH16, GH17, GH26, GH31, and GH52. Instead, *retaining* glycosynthases have been obtained only from hyperthermophilic exo- $\beta$ -glycosidases from family GH1 (for reviews see (Perugino et al. 2004) (Perugino et al. 2005), and (Hancock et al. 2006).

Despite the convenience of glycosynthase approach, GHs recalcitrant to become glycosynthases are not uncommon (Ducros et al. 2003) (Cobucci-Ponzano et al. 2003) (Di Lauro et al. 2008). In particular, while  $\beta$ -glycosynthase can be produced with a rather well established technology (Perugino et al. 2004), examples of  $\alpha$ -glycosynthases, for the synthesis, for instance, of  $\alpha$ -fucosylated,  $\alpha$ -mannosylated, and  $\alpha$ -galactosylated oligosaccharides with remarkable biomedical importance, are thus far limited only to enzymes from families GH29 (Cobucci-Ponzano et al. 2009), GH31 (Okuyama et al. 2002) and GH95 (Wada et al. 2008). Among these, two *retaining*  $\alpha$ -fucosynthases (namely the mutants SsD242S and TmD224G, of the  $\alpha$ -L-fucosidases of *S. solfataricus* and *T. maritima* respectively) from family GH29 have been described recently (Cobucci-Ponzano et al. 2003). Remarkably, these enzymes utilize  $\beta$ -fucosyl azide rather than  $\beta$ -fucosyl fluoride as *donor* substrate in *inverting*  $\alpha$ -fucosynthase reactions and they were not able to catalyse the synthesis using formate as external nucleophile (Figure 8.1)(Cobucci-Ponzano et al. 2009). Presumably,  $\beta$ -D-fucosyl formate was not stable enough to act as intermediate; in

contrast,  $\beta$ -FucN<sub>3</sub> had the right balance between reactivity and stability to allow product synthesis.



**Figure 8.1:** Catalytic mechanism of TmD224G  $\alpha$ -L-Fucosynthase (Cobucci-Ponzano et al. 2009)

Other special cases related to glycosynthases development, which follow atypical reaction mechanisms, are reported in the literature. In particular, the *retaining* endo- $\beta$ -N-acetyl-glucosaminidase from *Mucor hiemalis* (GH85), which follow a substrate-assisted catalysis, was converted into a glycosynthase by mutating the residue involved in the appropriate orientation of the substrate (Umekawa et al. 2008). Differently, two *inverting* glycosidases, namely the exo-oligoxylanase from *B. halodurans* (GH8), and the 1,2- $\alpha$ -L-fucosidase from *Bifidobacterium bifidum* (GH95), acted as glycosynthases once mutated either in the residues working as general base or holding a catalytic water (Honda et al. 2006; Wada et al. 2008).

The variety of strategies to convert glycosidase into a glycosynthase clearly explains why efforts in the characterization of catalytic mechanism of glycoside hydrolases and in developing new methods to improve glycosynthases are constantly required.

## **Aim of the thesis**

The importance of carbohydrates in several biological processes is directly mirrored in a wide number of biotechnological applications, based mainly on glycoside hydrolases (GHs), including conversion of agricultural byproducts in fermentable sugars for the bioethanol production (Waclawovsky et al. 2010), the use of these biocatalysts in the formulation of laundry detergents and in food industry (e.g. hydrolysis of lactose and preparation of HFCS) (Bommarius et al. 2004). Moreover, GHs are used also in several biomedical approaches, as the production of universal blood (Q. P. Liu et al. 2007), the enzyme replacement therapy for the treatment of lysosomal storage diseases (Phenix et al. 2010) and as alternative to the chemical synthesis of therapeutic biomolecules as heparin and galactooligosaccharides (Seeberger et al. 2007) (Schwab et al. 2010).

The development of a new class of enzyme, the glycosynthases (GS), obtained by mutating glycoside hydrolases, represent reliable alternative for the chemo-enzymatic synthesis of oligosaccharides. Here, the key of this approach is to cancel the hydrolytic activity of the enzyme by site-directed mutagenesis, but maintaining intact the structure of the active site. Therefore, by using substrate *donor* and certain reaction conditions, the engineered enzymes are able to synthesize products in quantitative yield.

This thesis is directed to the identification and characterization of glycosyl hydrolase for biotechnological applications and was divided in two different parts. The first part (Chapters I and II) is aimed to the application of glycosidases in the oligosaccharides synthesis. In particular, I addressed my work to the study and characterization of the catalytic mechanism of a  $\beta$ -galactosidase from the moderate thermophile *Alicyclobacillus acidocaldarius* (Aa $\beta$ gal) for the development of a new  $\beta$ -galactosynthase (Chapter I). Moreover I have characterized a new  $\alpha$ -galactosynthase from the hyperthermophile *Thermotoga maritima* (TmGalA D327G) (Chapter II) to validate the approach based on the use of  $\beta$ -glycosyl azide *donors* recently proposed (Cobucci-Ponzano et al. 2009).

The second part of this thesis (Chapters III and IV) is dedicated to the characterization in detail of two new glycoside hydrolases: a  $\beta$ -glycosidase SSO1353 and a  $\alpha$ -mannosidase (Ss $\alpha$ -Man) both from the hyperthermophilic crenarchaeon *Sulfolobus solfataricus*. The characterization of SSO1353 was directed to the study of the catalytic mechanism and in particular to the identification of nucleophile and acid/base residues. Differently, the study of Ss $\alpha$ -Man activity was directed to the analysis of its substrate specificity toward glycoconjugates and glycoproteins.

## **Experimental procedures**

## Bacterial Strains

***Escherichia coli* RB791** *lac I* p4000 (*lac I*<sup>q</sup>), *lac Z* p4008 (*lac L*8),  $\lambda$ -, IN( *rrn D* - *rrn E*)

***Escherichia coli* BL21(DE3)** *fhuA2* [*lon*] *ompT gal* ( $\lambda$  DE3) [*dcm*]  $\Delta$ *hsdS*

***Escherichia coli* BL21(DE3)RIL** F<sup>-</sup> *ompT hsdS*(r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>) *dcm*<sup>+</sup>Ter<sup>r</sup> *gal*  $\lambda$  (DE3) *endA Hte* (*argU ileY leuW Cam*<sup>r</sup>)

***Sulfolobus solfataricus* strain P2**

## Culture media

**LB (Luria-Bertani Broth) (1 liter):**

10 g NaCl,  
5 g yeast extract,  
10 g tryptone.

***S. solfataricus* minimal salts medium**

yeast extract (0.1%),  
casamino acids (0.1%),  
plus carbon source (0.1%)

## Chemicals

All commercially available substrates were purchased from Sigma and Carbosynth.

The  $\beta$ -GalN<sub>3</sub> was chemically synthesized, by Dr. E. Bedini of University of Naples "Federico II", from galactose in three steps (peracetylation, stereoselective anomeric azidation and deacetylation) and 85% overall yield according to a reported procedure (Györgydeák & Szilágyi, 1987)

## Detection of transgalactosylation activity by Thin Layer Chromatography (TLC)

A volume of 10-20  $\mu$ L of the reactions mixtures, after the time of incubation, was loaded and separated on a 20x20 silica gel 60 F254 TLC (Merck) using ethyl acetate-methanol-water (70:20:10) as eluant. Compounds were detected with UV light to determine the presence of UV-visible compounds, and by exposure to 4%  $\alpha$ -naphthol in 10% sulphuric acid in ethanol followed by heating (100°C) for the time necessary to signal the emergence of sugar.

## Site-directed mutagenesis

All mutants present in this thesis are obtained using the GeneTailor™ Site-directed Mutagenesis System kit (Invitrogen) and the primers listed in Table 1.2 using as the template plasmids obtained in the laboratory where I performed my thesis.

This kit relies on the inherent properties of two enzymes, DNA methylase and Mcr BC endonuclease as shown in the workflow diagram below:

1. Methylation of the plasmid DNA with a methylase;
2. Amplification of the plasmid in a mutagenesis reaction with two overlapping primers (forward, that contains the target mutation, and reverse). The product is linear, double stranded DNA containing the mutation.
3. Transform *E. coli* cells with the mutagenesis mixture. The host cell circularizes the linear mutated DNA while Mcr BC endonuclease digest *in vivo* the methylated template DNA, leaving only unmethylated, mutated product.

Name	Sequence
E313G mut	CTCAAGAAGCCATTCTGCTCAT <u>GGG</u> GTCCACGCCGAG
E313G rev	CATGAGCAGAAATGGCTTCTTGAGGATGGCGCGG
E157G mut	GTGATCGGCTGGCACGTGTCGAAC <u>GGG</u> TACGGCGGCG
E157G rev	GTTCGACACGTGCCAGCCGATCACGCCCGGATGATG
D276G mut2	GCTCGACGTCATCTCGTGG <u>GCG</u> AGCTATCCGC
D276G rev2	CCCACGAGATGACGTCGAGCACGTGCGGGAAG
E361G mut as	ACCACGGCGCCGTGGAAC <u>TTT</u> CATACGATCCGC
E157S mut as	CAGTGGCATTGCGCCCGTAC <u>GCT</u> GTTTCGACACGT
E157Q mut as	CAGTGGCATTGCGCCCGTAC <u>CTG</u> GTTTCGACACGT
E361G rev s	AAAGTTCCACGGCGCCGTGGTCGATCATGTCG
E313S mut s	CTCAAGAAGCCATTCTGCTCATG <u>TGG</u> TCCACGCCGAG
E313D mut s	CTCAAGAAGCCATTCTCCTCATG <u>GACT</u> CCACGCCGAG

**Table 1.2:** Site Directed Mutagenesis oligos of A $\beta$ -gal. The mismatched bases are underlined.

## Characterization of A $\beta$ -gal wild type and mutants

### Expression and purification of recombinant A $\beta$ -gal wild type and mutants

The mutants of *A. acidocaldarius*  $\beta$ -galactosidase (A $\beta$ -gal) are obtained from the template pGEX-2TK (GE Healthcare) in which was previously cloned the wild type enzyme (Di Lauro et al. 2008).

The synthetic oligonucleotides are from PRIMM (Italy), and listed in Table 3. All mutant genes were sequenced to ascertain the presence of the desired mutation.

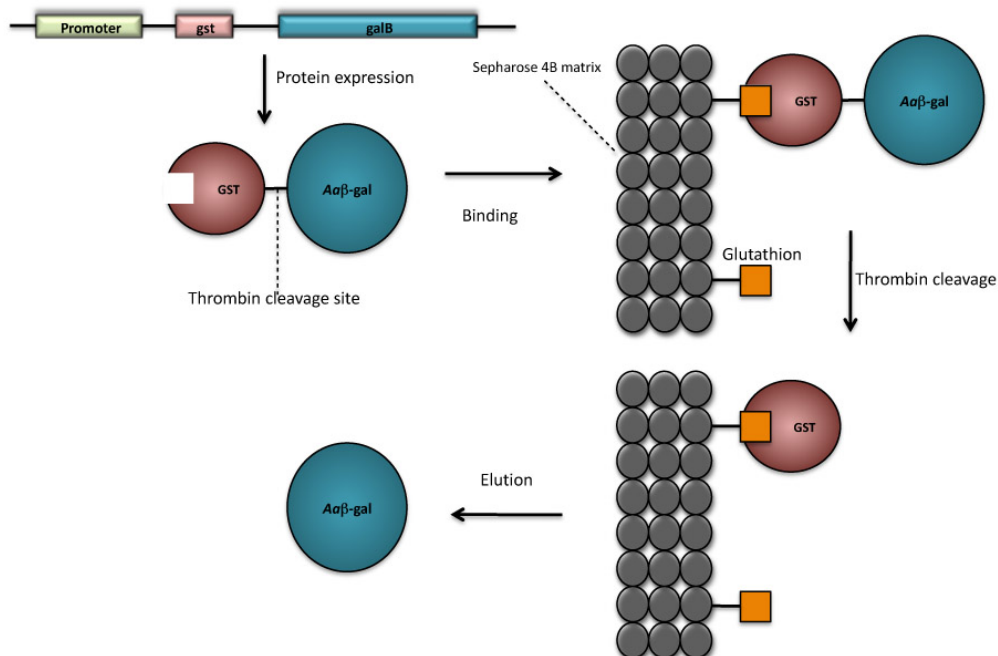
The mutants and wild type of A $\beta$ -gal from *A. acidocaldarius* were expressed as fusion proteins with Glutathione-S-transferase (GST). The plasmid containing the gene, pGex-A $\beta$ gal (amp<sup>r</sup>), was used to transform *E. coli* RB791 cells. They were grown in 2 Lt of LB broth, with ampicillin 50  $\mu$ g ml<sup>-1</sup>, at 37°C. Expression of the gene was induced by the addition of 0.5 mM IPTG when the culture reached an OD<sub>600nm</sub> of 1.0.

Growth allowed to proceed for 16 h and cells were harvested by centrifugation at 5,000 x g. The resulting cell pellet was resuspended in PBS 1X (150 mM NaCl, 20 mM phosphate buffer pH 7.3) with 1% TRYTON X-100 in a ratio1:3 (w:v) and incubated for 60 min at 37°C with 20 mg of lysozyme and 10 mg of Benzonase® (Novagen) to lysate cells and hydrolyze DNA, respectively.

Subsequently, the sample was homogenized by French cell pressure treatment. After centrifugation for 20 min at 30,000 x g the crude extract was purified by affinity



chromatography on Glutathione Sepharose™ 4B (Amersham). The active pool is then subjected to thrombin treatment on the resin in order to recover Aaβ-gal wild type and mutants free of the fused GST polypeptide (Figure 1.2). This purification procedure was achieved with a matrix that is dedicated only to the purification of each specific mutant, in order to exclude contamination by the wild type enzyme.



**Figure 1.2:** Schematic diagram of GST purification method for Aaβ-gal

The enzymes resulted >95% pure by SDS-PAGE. Protein concentration was determined with the method of Bradford (Bradford 1976), by using bovine serum albumin as standard. The samples stored at 4°C in PBS buffer were stable for several months.

### Enzymatic characterization of Aaβ-gal wild type and mutants

The standard assay for the β-galactosidase activity was performed in 50 mM sodium citrate buffer at pH 5.5 at 65°C on 2NP-β-D-Gal 20 mM. Typically, in each assay we used 0.05-2 μg of enzyme in the final volume of 1.0 ml.

Kinetic constants of Aaβ-gal wild type and mutants, on aryl-glycosides, were measured at standard conditions at 65°C by using concentrations of substrate ranging between 0.01 and 40 mM as reported in Table 2.2. The  $\epsilon_{mM}$  extinction coefficients for 2-nitrophenol under standard conditions and 65°C was  $1.14 \text{ mM}^{-1} \text{ cm}^{-1}$ .

Enzyme	[2NP- $\beta$ -D-Gal] mM
wt	between 0.1 and 28.0
E157G	between 0.05 and 5.0
D276G	between 0.01 and 20.0
E313G	between 0.05 and 40.0
E313D	between 0.1 and 40.0
E361G	between 0.1 and 30.0

**Table 2.2:** Concentration of 2NP- $\beta$ -D-Gal used for kinetic analysis

### Nano-HPLC-ESI-MS/MS Experiments

Mutant Aa $\beta$ -gal Glu313Gly (27  $\mu$ g, 0.35 nmol) was incubated for 2 hours with 2.9 mM 2,4-dinitrophenyl- $\beta$ -D-2-deoxy-2-fluoro-galactopyranoside (2,4DNP-2F-Gal) (obtained in collaboration with Dr. S.G. Withers of University of British Columbia) at 1:10<sup>3</sup> and 1:10<sup>5</sup> enzyme/inhibitor ratio in 50 mM sodium citrate buffer, pH 5.5 at 45°C. An identical mixture containing all the reagents with the exception of the inhibitor was prepared as control.

Protein (0.12  $\mu$ g/ $\mu$ l) was enzymatically digested in the acidified inhibition buffer (formic acid to 5% (v/v) final concentration, pH 2) using pepsin from porcine stomach mucosa (3,260 units/mg, Sigma-Aldrich) at 1:20 enzyme to substrate ratio at 37°C for 30 min. Resulting peptide mixtures (10  $\mu$ l) were loaded, purified, and concentrated on a monolithic trap column (200  $\mu$ m inner diameter  $\times$  5 mm, LCPackings, Sunnyvale, CA) at 25  $\mu$ l min<sup>-1</sup> flow rate and separated by nanoflow reverse-phase chromatography on a PS-DVB monolithic column (200  $\mu$ m inner diameter  $\times$  5 cm, LCPackings) at 300 nl min<sup>-1</sup> using an UltiMateTM 3000 HPLC (Dionex, Sunnyvale, CA). The following solvents and gradient conditions were used: solvent A: 2% acetonitrile in 0.1% formic acid and 0.025% trifluoroacetic acid, solvent B: 98% acetonitrile in 0.1% formic acid and 0.025% trifluoroacetic acid, gradient: 5% B for 5', 5–50% B in 60 min, 50–98% B in 6 s. Eluting peptides were directly analyzed by nano-ESI-MS in positive ion mode using information-dependent acquisition (IDA). The two most abundant multiply charged ions were automatically selected and subjected for collision induced dissociation experiments. Nitrogen was used as collision gas. Tandem mass spectra were analyzed by manual inspection and by the use of MASCOT Server (version 2.2). Peak lists for Mascot containing all acquired MS/MS spectra were generated by Analyst QS 2.0 software using the default parameters. Mascot was set up to search database containing a single protein (Aa $\beta$ -gal Glu313Gly) sequence and was run with a fragment ion mass tolerance of 0.1 Da and a parent ion tolerance of 50 ppm. MS/MS ion score cut-off was set to 10. No enzyme was specified. 2F-Gal was defined as variable modification in Mascot searches.

### **Characterization of TmGalA wild type and mutant**

#### Expression and purification of TmGalA wild type and Asp327Gly mutant

TmGalA wild type and mutant Asp327Gly were purified from 2 liter culture of *E. coli* BL21(DE3)/pET-AGT1, in LB medium supplemented with 100  $\mu$ g ml<sup>-1</sup> ampicillin. For

induction of the P<sub>T7</sub> promoter, 0.5 mM IPTG is added during the exponential growth phase at an OD<sub>600nm</sub> of 0.8. Growth was allowed to proceed for 16 h, and cells were harvested by centrifugation at 5,000 x g. The resulting cell pellet was thawed, resuspended in 1 ml g<sup>-1</sup> cells of 20 mM Tris-HCl pH 8 and homogenized by French cell pressure treatment. After centrifugation for 30 min at 10,000 x g, the crude extract was heat-fractionated for 20 min 75°C in order to denature heat-labile host *E. coli* proteins. The supernatant obtained after heat-fractionations, was applied to a HiLoad 16/10 Q-Sepharose High performance (Amersham Biotech), which had been equilibrated with 20 mM Tris-HCl, pH 8. Proteins bound to the column are eluted with a linear NaCl gradient (0 to 1.0 M) in 3 column volumes. Active fraction were pooled and brought to 5 M NaCl by slowly adding solid NaCl. The sample is then loaded onto a Phenyl-sepharose HP 26/10 (Amersham Biotech), and eluted with a linear NaCl gradient (5 to 0 M) in 6 column volumes in 20 mM Tris-HCl pH 8. Pooled fractions containing TmGalA were dialyzed against 20 mM sodium phosphate buffer, pH 7.3, 150 mM NaCl and successively concentrated by ultrafiltration on an Amicon YM30 membrane (cut off 30,000 Da).

#### Enzymatic characterization of TmGalA Asp327Gly

The  $\alpha$ -galactosidase activity assays, the steady state kinetic studies, and the azide rescue reactions were performed as previously reported (1  $\mu$ g of wild type and 10  $\mu$ g of Asp327Gly mutant), but at 65°C and without added bovine serum albumin (Comfort et al. 2007). Suitable blanks, containing all the reagents with the exception of enzyme, were always used to take into account the negligible spontaneous hydrolysis of the substrates. One enzymatic unit is defined as the amount of enzyme catalyzing the conversion of one  $\mu$ mole of substrate into product in one min, at the indicated conditions. All kinetic data were calculated as the average of at least two experiments and were plotted and refined with the program *GraFit* (Leatherbarrow 1992).

#### Transgalactosylation trials of TmGalA Asp327Gly

The glycosynthetic reactions were performed by incubating Asp327Gly (10  $\mu$ g) for 16 h at 65°C in 0.1 ml of 50 mM sodium acetate buffer pH 5.0 at the indicated concentrations of  $\beta$ -GalN<sub>3</sub> (donor) and the suitable acceptor. Blank mixtures without enzyme were always prepared. The products distribution was evaluated by thin layer chromatography (TLC) as reported above. The transgalactosylation efficiency of Asp327Gly mutant was measured by use of a High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) equipped with a PA200 column (Dionex, USA). Samples were eluted with 20 mM NaOH at a flowrate of 0.5 ml min<sup>-1</sup>. To measure the total amount of galactose enzymatically transferred, 1/10 of the reaction mixtures were incubated for 90 min at 65°C in the presence of 5.2  $\mu$ g TmGalA wild type. The efficiency of the transgalactosylation reaction was calculated as: total amount of galactose transferred - moles of galactose transferred to water / total amount of galactose transferred x 100.

#### Characterization of the galactosylated oligosaccharides obtained by TmGalA Asp327Gly

The chemical characterization, of different transgalactosylation products, was performed by the group of Prof. M.M. Corsaro of the University of Naples "Federico

II". The glycosynthetic reactions were performed at the same conditions described above in a total volume of 2 ml. The transgalactosylation products were purified by reverse phase chromatography (Polar-RP 80A, Phenomenex, 4  $\mu$ , 250 x 10 mm) on an Agilent HPLC instrument 1100 series and revealed by UV at 220 nm. Samples were eluted using the following conditions: 40% of methanol in water for reaction I; 40% of methanol in water for 5 min, 40% to 50% in 30 min, 50% for 15 min for reaction II; 30% methanol in water for 5 min, 30% to 50% in 30 min, 50% for 15 min for reactions III and IV. For all the disaccharides isolated except for the product **3**, the structural determination was obtained by  $^1\text{H}$  mono-dimensional and homonuclear ( $^1\text{H}$ ,  $^1\text{H}$ ) and heteronuclear ( $^1\text{H}$ ,  $^{13}\text{C}$ ) two-dimensional NMR experiments and by methylation analysis. For the detailed characterization of different products see Cobucci-Ponzano et al. doi: 10.1093/glycob/cwq177.

## Identification of the catalytic residues of SSO1353

### Expression and purification of SSO1353 wild type and mutants

The plasmid pET29a (Novagene) containing *sso1353* wild type gene (pET1353), and its mutants E335G, D406G, D458G and D462G, were prepared previously by cloning and site directed mutagenesis, respectively (Cobucci-Ponzano et al. 2010).

*E. coli* BL21(DE3)Ril/pET1353 wild type and mutants were grown in 2 liters of LB at 37°C supplemented with kanamycin (50  $\mu\text{g ml}^{-1}$ ) and chloramphenicol (30  $\mu\text{g ml}^{-1}$ ). Gene expression was induced by the addition of 0.5 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) when the culture reached an optical density of 1.0 at 600 nm. Growth was allowed to proceed for 16 h, cells were harvested by centrifugation at 5,000  $\times g$  and stored at -20°C. The resulting cell pellet was thawed, resuspended in 3 ml  $\text{g}^{-1}$  cells of 20 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, 1% (v/v) Triton X-100 and homogenized by French cell pressure treatment.

After centrifugation for 30 min at 10,000  $\times g$ , the crude extract was incubated with Benzonase® (Novagen) for 1 h at room temperature and then heat-fractionated for 30 min at 55 and 75°C and for 20 min at 85°C. The supernatant obtained after heat-fractionations, equilibrated in 1 M ammonium sulphate, was applied to a HiLoad 26/10 Phenyl Sepharose High performance (Amersham Biotech), which had been equilibrated with 20 mM sodium phosphate buffer, pH 7.3, 1 M ammonium sulphate. After washing with 2 column volumes with the loading buffer, the protein was eluted with a linear gradient of water at a flow rate of 3 ml  $\text{min}^{-1}$ ; the protein eluted in 100% water. Active fractions were pooled, equilibrated in 20 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl and concentrated by ultrafiltration on an Amicon YM30 membrane (cut off 30,000 Da). For the wild type enzyme, after concentration, the sample was loaded onto a HiLoad 26/60 Superdex 200 prep grade column (Amersham Biotech). Active fractions were pooled and concentrated; protein concentration was determined with the method of Bradford (Bradford, 1976).

The SSO1353 wild type and mutants were 95% pure by SDS-PAGE and were stored at 4°C.

### Standard assay of SSO1353 and mutants

The standard assay for SSO1353 wild type and mutants activity was performed in 50 mM sodium citrate buffer at pH 5.5 at 65°C on 2NP-Glc. Typically, in each assay we used 1-10 µg of enzyme in the final volume of 1.0 ml. The  $\epsilon_m$ M extinction coefficients at 405 nm for 2-nitrophenol under standard conditions and 65°C was 1.1 mM<sup>-1</sup> cm<sup>-1</sup>. One unit of enzyme activity was defined as the amount of enzyme catalyzing the hydrolysis of 1 µmol of substrate in 1 min at the conditions described.

### Chemical rescue activity of SSO1353 D462G mutant

The chemically rescued activity of the D462G mutant was measured in 50 mM sodium citrate buffer at pH 5.5 on 40 mM 2NP-Glc at 65°C. The assay mixture was supplemented with concentrations of sodium azide as external nucleophile between 0 and 1.0 M. In all of the assays, spontaneous hydrolysis of the substrate was subtracted by using appropriate blank mixtures without the enzyme.

### Analysis of the products obtained by chemical rescue

Aliquots of the reaction mixtures were analyzed on a silica gel 60 F<sub>254</sub> TLC (Merck) by using ethyl acetate/methanol/water (70:20:10 v/v) as eluant and were detected by exposure to 4%  $\alpha$ -naphthol in 10% sulphuric acid in ethanol followed by charring.

The  $\beta$ -D-glucosyl azide ( $\beta$ -GlcN<sub>3</sub>) isolated from the enzymatic reaction mixture of the D462G mutant was identified by <sup>1</sup>H and <sup>13</sup>C NMR spectroscopy. For the detailed NMR characterization see Cobucci-Ponzano et al. 2010.

### Inhibition of SSO1353 wild type

The effect of the inhibitor 2,4-dinitrophenyl- $\beta$ -D-2-deoxy-2-fluoro-glucopyranoside (2,4DNP-2F-Glc) (Sigma) was analyzed as previously performed by Shaikh and co-workers (Shaikh et al. 2007).

Wild type SSO1353 (0.1 µg/µl) was incubated at 45°C in mixtures containing 0.3, 3.0, 7.0, and 18.0 mM concentrations of inhibitor and 50 mM sodium citrate buffer, pH 5.5. An identical mixture containing all the reagents with the exception of the enzyme was prepared as control. At time intervals, aliquots from the two mixtures were withdrawn and used to measure the enzymatic activity and as blank, respectively. Assays were performed on 60 mM 2NP-Glc in standard conditions. Initial rates at each time point were elaborated, to measure the inactivation parameters K<sub>i</sub> and k<sub>i</sub> with the software GraFit (Leatherbarrow 1992).

### Nano-ESI-MS of Intact Protein Samples

Samples were analyzed using a triple quadrupole time of flight instrument (QSTAR Elite, Applied Biosystems) equipped with a nanoflow electrospray ion source.

Pulled silica capillary (170 µm outer diameter/100 µm inner diameter, tip 30 µm inner diameter) was used as nanoflow tip. For the analysis of intact proteins, 4 µg of samples were purified using ZipTip C4 (Millipore). Proteins were eluted by 50% acetonitrile and 0.1% formic acid.

Purified proteins (10  $\mu$ M) were loaded into the ion source at 300 nl min<sup>-1</sup> flow rate using a syringe pump. Single-stage ESI mass spectra were acquired in the range of m/z 300–2000. For protein molecular mass determination three independent measurements were performed. The expected mass error on the average molecular mass of intact proteins was about  $\pm 0.01\%$ .

For data acquisition and Bayesian protein reconstruction the Analyst QS 2.0 software (Applied Biosystems, Foster City, CA/Toronto, Canada) was used.

#### Nano-HPLC-ESI-MS/MS Experiments

Wild-type SSO1353 (22  $\mu$ g, 0.3 nmol) was incubated with 2.9 mM 2,4-dinitrophenyl- $\beta$ -D-2-deoxy-2-fluoro-glucopyranoside (2,4DNP-2F-Glc) (Sigma) at 1:1000 enzyme/inhibitor ratio in 50 mM sodium citrate buffer, pH 5.5 at 45°C. An identical mixture containing all the reagents with the exception of the inhibitor was prepared as control. At time intervals, aliquots from the two mixtures were withdrawn and assayed on 60 mM 2NP-Glc in standard conditions.

Samples (0.086  $\mu$ g/ $\mu$ l) were enzymatically digested in the acidified inhibition buffer (formic acid to 5% (v/v) final concentration, pH 2) using pepsin from porcine stomach mucosa (3,260 units/mg, Sigma-Aldrich) at 1:20 enzyme to substrate ratio at 37°C for 30 min. Resulting peptide mixtures (5  $\mu$ l) were loaded, purified, and concentrated on a monolithic trap column (200  $\mu$ m inner diameter  $\times$  5 mm, LCPackings, Sunnyvale, CA) at 25  $\mu$ l min<sup>-1</sup> flow rate and separated by nanoflow reverse-phase chromatography on a PS-DVB monolithic column (200  $\mu$ m inner diameter  $\times$  5 cm, LCPackings) at 300 nl min<sup>-1</sup> using an UltiMate<sup>TM</sup> 3000 HPLC (Dionex, Sunnyvale, CA). The following solvents and gradient conditions were used: solvent A: 2% acetonitrile in 0.1% formic acid and 0.025% trifluoroacetic acid, solvent B: 98% acetonitrile in 0.1% formic acid and 0.025% trifluoroacetic acid, gradient: 5–50% B in 40 min, 50–98% B in 6 s. Eluting peptides were directly analyzed by nano-ESI-MS in positive ion mode using information-dependent acquisition (IDA). The two most abundant multiply charged ions were automatically selected and subjected for collision induced dissociation experiments. Nitrogen was used as collision gas. Tandem mass spectra were analyzed by manual inspection and by the use of Mascot Server (version 2.2). Peak lists for Mascot containing all acquired MS/MS spectra were generated by Analyst QS 2.0 software using the default parameters. Mascot was set up to search database containing a single protein (SSO1353) sequence extracted from NCBI nr and was run with a fragment ion mass tolerance of 0.1 Da and a parent ion tolerance of 50 ppm. MS/MS ion score cut-off was set to 10. No enzyme was specified. 2F-Glc was defined as variable modification in Mascot searches. Three independent inhibition experiments were performed and on each resulting samples two analytical measurements were run.

All the mass-spectrometry experiments were performed in collaboration with Dr. G. Pocsfalvi of Institute of Protein Biochemistry – Italian National Research Council.

## Characterization of the substrate specificity of nSs $\alpha$ -man on mannosylated glycans

### Purification of native $\alpha$ -mannosidase from *S. solfataricus* (nSs $\alpha$ -man)

*S. solfataricus* cells, strain P2, were grown at 80°C, pH 3.0 in a minimal salts medium supplemented with yeast extract (0.1%), sucrose (0.1%), and casaminoacids (0.1%). Growth was monitored spectrophotometrically at 600 nm. When the culture reached an A<sub>600</sub> of 0.6 optical densities, cells were harvested by centrifugation at 5000 × g.

The resulting cell pellet was thawed, resuspended in 3 ml g<sup>-1</sup> cells of 20 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, 1% (v/v) Triton X-100. Cells were lysated by three cycles of freeze thawing (5 min at -70°C; 5 min 37°C), and centrifuged for 30 min at 10,000 × g. The crude extract was then applied at a flow rate of 2.5 ml min<sup>-1</sup> on a High Load 16/10 Q-Sepharose High Performance column (Amersham Biotech) equilibrated in 20 mM phosphate buffer, pH 7.3. At these conditions, the protein did not bind to the column.

Active fractions were pooled, equilibrated in 1 M ammonium sulphate, and applied to a HiLoad 26/10 Phenyl Sepharose High performance (Amersham Biotech), which was equilibrated with 20 mM sodium phosphate buffer, pH 7.3; 1 M ammonium sulphate.

After washing with 1-column volumes with the loading buffer, the protein was eluted with a two-step gradient of water (0–80%, 2 column volumes; 80–100%, 3 volumes; 100% 2 volumes) at a flow rate of 3 ml min<sup>-1</sup>; the protein eluted in 90% water. Active fractions were pooled, dialyzed against 20 mM sodium phosphate buffer, pH 7.3, and concentrated by ultrafiltration on an Amicon YM30 membrane (cut off 30,000 Da).

After concentration, the sample was loaded on High Load 16/10 Q-Sepharose High Performance column (Amersham Biotech) equilibrated in 20 mM phosphate buffer, pH 7.3. After washing with 2 column volumes with the loading buffer, the protein was eluted with a two-step gradient using 20 mM phosphate buffer, pH 7.3; 1 M NaCl (0–25%, 4 column volumes; 25–100% 1 volume; 100% 2 volumes), at a flow rate of 3 ml min<sup>-1</sup>. The protein eluted in 10% 20 mM phosphate buffer, pH 7.3; 1 M NaCl.

Active fractions were pooled, concentrated, and loaded on a Superdex 200 HR 10/30 gel filtration column for FPLC, equilibrated in 20 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl at a flow rate of 1 ml min<sup>-1</sup>. Active fractions were pooled and concentrated; protein concentration was determined with the method of Bradford (Bradford 1976). The native enzyme, named nSs $\alpha$ -man, was 95% pure by SDS-PAGE and was stored at 4°C.

### Expression and purification of rSs $\alpha$ -man

The recombinant  $\alpha$ -mannosidase (rSs $\alpha$ -man) was previously cloned in pGEX-2TK plasmid to obtain pGEX-man (Cobucci-Ponzano et al. 2010).

rSs $\alpha$ -Man was expressed in pGEX-MAN/RB791 *E. coli* cells without induction with IPTG. The recombinant enzyme was purified exploiting the GST-tag and the thrombin cleavage onto the matrix as described in *Experimental procedures I*.

The contaminating proteins were eliminated by two subsequent incubations for 20 min at 70°C and 80°C, each followed by a centrifugation at 10,000 × g for 20 min at 4°C and a Superdex 200 HR 10/30 gel filtration column for FPLC, performed as described above. After this procedure rSsα-man, was more than 95% pure by SDS-PAGE.

#### Enzymatic assays

The standard assay for the α-mannosidase activity was performed in 50 mM sodium phosphate buffer, pH 6.5 at 65°C. 4NP-α-D-Man was used at concentrations 5 mM in the final volume of 0.2 ml.

Typically, in each assay we used about 0.9–3 µg (8–27 pmoles) of enzyme. At appropriate times (typically from 1 to 10 min) reaction mixture was transferred in 0.8 ml of an iced solution of 1 M Na<sub>2</sub>CO<sub>3</sub> and then the absorbance was determined at 420 nm ( $\epsilon_{mM} = 17.2 \text{ mM}^{-1} \text{ cm}^{-1}$ ). One enzymatic unit is defined as the amount of enzyme catalysing the conversion of 1 µmole of substrate into product in 1 min, at the indicated conditions.

#### Reactivation of the mutant Ssα-man Asp338Gly

The reactivation of the D338G mutant was measured in 50 mM sodium phosphate buffer, 0.1 mM ZnCl<sub>2</sub> at pH 6.5 on 2.5 mM 4NP-α-D-Man at 65°C. The assay mixture was supplemented with concentrations of sodium formate as external nucleophile between 0.5 and 2.0 M.

#### Activity of Ssα-man on N-linked oligosaccharides

The activity of nSsα-man on Man<sub>3</sub>GlcNAc<sub>2</sub> and Man<sub>7</sub>GlcNAc<sub>2</sub> oligosaccharides was measured by incubating the native enzyme (10.6 µg, 96 pmoles) in 50 mM sodium acetate buffer, pH 5.0 and 2 mM ZnCl<sub>2</sub> for 16 h at 50°C.

Aliquots were withdrawn at different times and stored in ice. Reaction products were analysed by High-Performance Anion-Exchange chromatography with Pulsed Amperometric Detection (HPAE–PAD) equipped with a PA200 column (Dionex, USA). Runs were performed in 50 mM NaOH with a linear gradient of 1–8 mM sodium acetate, 0.5 ml min<sup>-1</sup> for 60 min.

#### De-mannosylation analysis of Rnase B

Rnase B (Sigma–Aldrich) 1.2 nmoles (12 µM) were incubated with 96 pmoles (96 µM corresponding to 0.03 units) of nSsα-man in 10 mM sodium acetate buffer, pH 5.0 and 2 mM ZnCl<sub>2</sub>, for 16 h at 50°C. Blank sample was prepared in the same way but in the absence of nSsα-man. The reaction mixture was desalted by using ZipTip C4 (Millipore). Proteins were eluted in 10 µL 50% acetonitrile 0.1% (v/v) trifluoroacetic acid and spotted (1 µL) on a normal phase ProteinChip array (NP20; Bio-Rad) in triplicates. 1 µL of sinapinic acid (Bruker Daltonics) 5 mg ml<sup>-1</sup> in 50% acetonitrile 0.1% (v/v) trifluoroacetic acid was applied on the spots and let them air-dry.

A SELDI-TOF (Bio-Rad) mass spectrometer was used to measure the change in the molecular mass distribution of the various known glycoforms of Rnase B upon mannosidase treatment. Data were collected by 400 Hz sampling rate in the mass



range of  $m/z$  1000–40,000 and signal optimization range of  $m/z$  15,000. A total of 710 laser shots were taken for an averaged mass spectrum. Three independent measurements were performed on each spot. Instrument was calibrated externally with calmix-3 protein standard (Applied Biosystems).

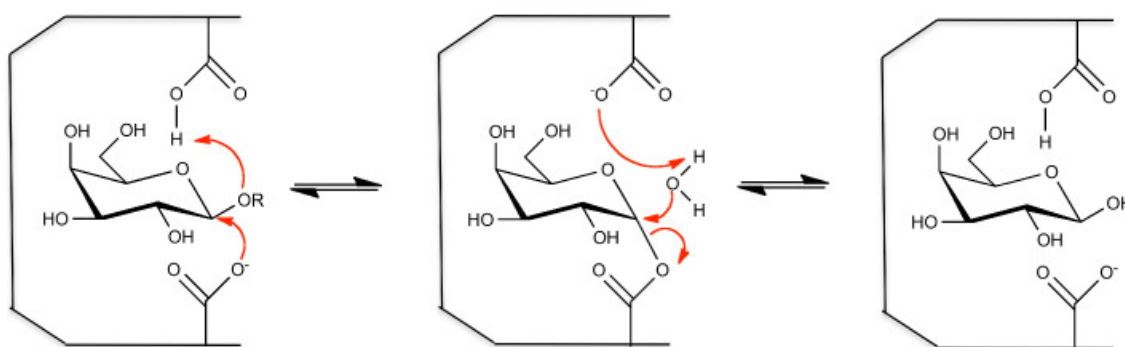
The samples prepared in the same conditions described above were also analysed by HPAE–PAD equipped with a PA200 column. Runs were performed in 100 mM NaOH,  $0.5\text{ ml min}^{-1}$  for 20 min.

# **Chapter I: A $\beta$ -galactosidase from *Alicyclobacillus acidocaldarius***

## **Introduction**

## The $\beta$ -Galactosidases

$\beta$ -Galactosidases (EC 3.2.1.23) are ubiquitous enzymes found in all three domains of life that are well known as gene reporters in molecular and cell biology and in the conversion of lactose for nutritional applications (Siso 1996). At present, an inspection of CAZy classification shows that  $\beta$ -galactosidases activities are grouped in glycoside hydrolase (GH) families GH1, 2, 35, and 42, all including *retaining* enzymes (Figure 1.3).



**Figure 1.3:** Catalytic mechanism of a retaining  $\beta$ -galactosidase

GH1 family groups, in addition to  $\beta$ -galactosidase (EC 3.2.1.23) and 6-phospho- $\beta$ -galactosidases (EC 3.2.1.85), also  $\beta$ -glucosidases (EC 3.2.1.21),  $\beta$ -mannosidases (EC 3.2.1.25),  $\beta$ -D-fucosidases (EC 3.2.1.38) and several other enzymatic activities. Many enzymes of this family have been biochemically characterized in detail and all showed wide substrate specificity (<http://www.cazy.org/GH1.html>).

The most common activities for glycoside hydrolases of GH2 family include  $\beta$ -galactosidases (EC 3.2.1.23),  $\beta$ -glucuronidases (EC 3.2.1.31),  $\beta$ -mannosidases (EC 3.2.1.25), exo- $\beta$ -glucosaminidases (EC 3.2.1.165) and, in plants, a mannosylglycoprotein  $\beta$ -mannosidase (EC 3.2.1.152). Enzymes from GH2 are found across a broad spectrum of life forms, but are most abundant in bacteria. The most studied enzyme in this family is the *E. coli*  $\beta$ -galactosidase (lacZ), a component of the lac operon while the principal enzyme of medical interest is the lysosomal  $\beta$ -glucuronidase whose deficiency leads to Sly syndrome (Sly et al. 1973). The only other human GH2 enzyme is the lysosomal  $\beta$ -mannosidase.

Glycoside hydrolases belonging to family GH35 are  $\beta$ -galactosidase (EC 3.2.1.23) and exo- $\beta$ -glucosaminidase (EC 3.2.1.165). This family includes multiple genes in various plant species, as *Arabidopsis*, tomato, papaya, apple, vigna, and barley and seems to play a key role in pit membrane modification and hydrolysis of primary walls, by degrading galactan side chains of rhamnogalacturanan I, or xyloglucan (Tanthanuch et al. 2008).

### The $\beta$ -galactosidases of GH42 family

The most common activity in family GH42 family is  $\beta$ -galactosidase (EC 3.2.1.23) followed by  $\alpha$ -D-arabinosidases (EC 3.2.1.55) and  $\beta$ -D-fucosidases (EC 3.2.1.38) suggesting a strict specificity for axial C4-OH groups (Kosugi et al. 2002) (Di Lauro et al. 2008). GH42 enzymes are active on lactose and are able to promote transgalactosylation reactions with production of galactooligosaccharides (Møller et al. 2001). They have been identified only in unicellular organisms, mainly from prokaryotes, with a few examples from archaea and fungi. Several GH42 enzymes are extracted from diverse habitats where lactose would not be present; however, they are very active on galactooligosaccharides and galactans (Van Laere et al. 2000), suggesting that, possibly, they are involved *in vivo* in plant cell wall degradation. This function could be performed in cooperation with family GH53 galactanases, often encoded from genes clustered near GH42 genes (Shipkowski et al. 2006).

On the basis of hydrophobic cluster analysis (HCA), GH42 enzymes are classified in the clan GH-A, which group 18 families showing the typical TIM barrel fold and following the *retaining* reaction mechanism (Henrissat et al. 1995).

At present, only the 3D-structure of the  $\beta$ -galactosidase from *T. thermophilus* A4 (A4- $\beta$ -Gal) is available in GH42, solved at 1.6 Å and 2.2 Å resolution as either free and galactose-bound, (PDB ID: 1KWG, 1KWK, respectively) (Hidaka et al. 2002). The inspection of the latter suggested that Glu312 and Glu141 could be the nucleophile and the acid/base of the reaction, respectively.

The catalytic nucleophile was directly identified for the first time as Glu295 in the *B. subtilis* YesZ  $\beta$ -galactosidase through the use of a mechanism-based inhibitor and subsequent peptide mapping (see the Introduction of this thesis for a detailed description of this method) (Shaikh et al. 2007). At the time at which this PhD program started the general acid/base catalyst was unknown. The first report is that described here on the  $\beta$ -galactosidase from *A. acidocaldarius* (Di Lauro et al. 2008).

### Biotechnological applications of $\beta$ -galactosidases

$\beta$ -Galactosidases are largely exploited in the food industry for their capacity to hydrolyse lactose.

Adult-type hypolactasia is determined by a genetically programmed reduction in lactase activity at the intestinal brush border. The incidence of lactose maldigestion ranges from 11% to 60% in Europe and this condition can cause gastro-intestinal symptoms such as abdominal pain, bloating, flatulence and diarrhoea. Low lactose foods are commercially available, obtained by pre-treatment with recombinant  $\beta$ -galactosidase. In addition enzyme-replacement therapy (ERT) with microbial exogenous lactase (obtained from yeasts or fungi) administered in a solid form (capsules or tablets), allow million people with lactose intolerance to eating a balanced diet (Montalto et al. 2006).

Another important disaccharide present in the marketed milk, lactulose (4-O- $\beta$ -D-galactopyranosyl-D-fructofuranose), which is formed during milk heat treatment, has been proposed by the International Dairy Federation (IDF) and by the European Commission as analytical index to distinguish Ultra High Temperature (UHT) milk

from pasteurized milk. Moreover, lactulose is used in infant formula, due to its ability to promote the proliferation of *Lactobacillus bifidus*, as well as for prevention and treatment of chronic constipation, portal systemic encephalopathy and other intestinal or hepatic disorders. There are several analytical methods for detection of lactulose, namely, gas chromatography, liquid chromatography, and enzymatic kits based on spectrophotometric or amperometric detection (A. Amine et al. 2000). More, lactulose biosensors are based on immobilized  $\beta$ -galactosidase and fructose dehydrogenase. This procedure is very simple and rapid and was adopted by small, medium and large dairy industries. However, the process of immobilization require drastic treatments that may denature mesophilic proteins. Therefore, optimization of lactulose biosensors with thermophilic  $\beta$ -galactosidases intrinsically stable to immobilization and reaction conditions, could allow a sensible reduction of costs (Campuzano et al. 2004).

Another biotechnological application of  $\beta$ -D-galactosidase of considerable interest is the synthesis of galacto-oligosaccharides (GOS). These compounds, mainly D-Gal- $\beta$ (1-3)-D-Gal (3-galactobiose), D-Gal- $\beta$ (1-6)-Lac (6'-galactosyl-lactose), D-Gal- $\beta$ (1-3)-D-Glc (3-galactosyl-glucose), D-Gal- $\beta$ (1-3)-Lac (3'-galactosyl-lactose), D-Gal- $\beta$ (1-6)-D-Glc (allolactose) and D-Gal- $\beta$ (1-6)-D-Gal (6-galactobiose) (Martinez-Villaluenga et al. 2008) belong, because of their indigestible nature, to the group of prebiotics. They are non-digestible food ingredients that beneficially affect the host by stimulating the growth and/or activity of colon bacteria. GOS naturally occurs in human milk, and commercial available products are broadly used in infant formula, biscuits, and food for critical ill (Van Laere et al. 2000). Large amount of studies on both infants and adults have shown that consumption of galacto-oligosaccharides resulted in a significant increase in *Bifidobacteria* (Piirainen et al. 2008).

The large demand for these important nutraceuticals requires the develop of new efficient and economically sustainable methods for their synthesis. In these regards,  $\beta$ -D-galactosidases from a great variety of bacterial and fungal species attracted interest for their ability in catalyzing transgalactosylation reaction for the formation of GOS (Møller et al. 2001). To test the possibility to use  $\beta$ -galactosidases from thermophilic microorganisms for these applications, we have isolated and characterized in detail an enzyme from the moderately thermophilic bacterium *Alicyclobacillus acidocaldarius*.

#### The $\beta$ -galactosidase from *A. acidocaldarius*

*Alicyclobacillus acidocaldarius*, isolated for the first time in Yellowstone National Park USA, is a thermoacidophilic gram-positive bacterium that grows optimally in strictly aerobic conditions at 60°C and pH 3-4 (Darland et al. 1971).

*Alicyclobacilli* have been isolated from different habitats including geothermal sites, submarine hot spring, but also soft drinks and heat-processed foods. It has been demonstrated that *A. acidocaldarius* can use as carbon energy source several sugars including L-arabinose, ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, L-rhamnose, D-turanose, mannitol, melibiose, cellobiose, lactose, maltose, sucrose, trehalose, tagatose and the polysaccharides cellulose, xylan, starch and glycogen (Goto et al. 2002). More recently, the sequencing of the genome of strain DSM 446 demonstrated that this organism is an interesting source of glycoside hydrolases (<http://www.cazy.org/b1056.html>).

In the lab where I performed my PhD, a  $\beta$ -D-galactosidase was purified from extracts of *A. acidocaldarius* and its gene was cloned and expressed in *E. coli*. The recombinant enzyme (Aa $\beta$ -gal) was characterized in detail (Di Lauro et al. 2008). Its activity vs pH on 2NP- $\beta$ -D-Gal substrate showed a flattened bell-shaped curve at 65 °C with an optimum at pH 5.5. No dependence by divalent metal ion cofactors was observed.

The enzyme showed similar affinity ( $K_M$ ) for different aryl-glycosides while the catalytic efficiency ( $k_{cat}/K_M$ ) was in the order 2NP- $\beta$ -D-Gal > 4NP- $\alpha$ -D-Ara > 2NP- $\beta$ -D-Fuc. No activity was observed on gluco- and xylosides showing C4-OH group in the equatorial position, indicating that Aa $\beta$ -gal was extremely specific for the axial C4-OH group of the galacto-, fuco-, and arabinoside substrates. For the same reason, cellobiose was not a substrate of the enzyme while lactose and lactulose were hydrolysed efficiently (Table 1.3)

Substrate	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ (mM)	$K_{cat}/K_M$ (s <sup>-1</sup> mM <sup>-1</sup> )
4NP- $\beta$ -D-Gal	2988 $\pm$ 78	4.9 $\pm$ 0.4	609
2NP- $\beta$ -D-Gal	2657 $\pm$ 86	5.5 $\pm$ 0.6	484
4NP- $\alpha$ -L-Ara	1740 $\pm$ 41	5.2 $\pm$ 0.4	332
2NP- $\beta$ -D-Fuc	1345 $\pm$ 43	7.3 $\pm$ 0.6	186
Lactose	212 $\pm$ 5	60.8 $\pm$ 5.5	3
Lactulose	380 $\pm$ 16	18.2 $\pm$ 3.0	21

**Table 1.3:** Steady state kinetic constants of the recombinant Aa $\beta$ -gal

It is worth mentioning the higher specificity constant toward lactulose if compared to that on lactose, which makes of the thermally stable Aa $\beta$ -gal a promising alternative to the enzyme from *Aspergillus oryzae*, which is currently used for the determination of lactulose in milk (Amine et al. 2000).

## **Chapter I**

### **Results and Discussion**

## Characterization of $\beta$ -galactosidase mutants from *A. acidocaldarius* for the oligosaccharides synthesis

### Identification of catalytically active residues in Aa $\beta$ -gal

The aminoacidic sequence of Aa $\beta$ -gal was analyzed through T-coffee multi-alignment with eight different sequences belonging to GH42 family showing an identity  $\geq 30\%$ : A4- $\beta$ -Gal from *T. thermophilus* A4, the enzyme from *T. thermophilus* T2, BgalA from *T. neapolitana*, YesZ from *B. subtilis*, BgaB from *B. circulans*, BgaB from *G. stearothermophilus*, BgalC from *T. maritima* and BgaH from *H. lucentense* (Figure 2.3).

		<b>E157</b>	
A.acidocaldarius	IGWHVSN <b>E</b> YG----GECHCPLCQDAFRKWLRKY-KTLDALNHAWWTFW	194	
Thermus.A4	AGFQTDN <b>E</b> YGCHDTVRCYCPCQEAFRGWLEARY-GTIEALNEAWGTAFW	182	
Thermus.T2	VGFQVDN <b>E</b> FGCHGTVRCYCPCNCREAFRGWLRKY-GTIDALNAAWGTVFW	182	
T.neapolitana	VGWQTDN <b>E</b> YGCHDTVRCYCPCCKAFQKWLERRYEGDIDKLNRAWGTVFW	183	
B.subtilis	IGWQLDN <b>E</b> FKCH-VAECMCETCLRLWHDWLKNRY-GVIERLNEAWGTDVW	185	
B.circulans	IGWHISN <b>E</b> FG----GDCHCDYCQDAFRGWVKNKY-GTDELNHSWWTTFW	195	
G.stearothermophilus	KMWHVNN <b>E</b> YACH-VSKCFECNCAVAFRWKWKERY-KTIDELNERWGTNFW	188	
T.maritima	VLWHVNN <b>E</b> YL----NYCYCDICRGKFQNLKEKY-GTDELNRRWNTFRW	187	
H.lucentense	AGWQTDN <b>E</b> FGCHETVTCYCEDCGEAFSEWLADRY-ESVADLNDAWGTTFW	183	
		:: .**:	* * * : * : * : * * *
			<b>D276</b>
A.acidocaldarius	HEIEPLKQVNPPLPVTTNFMG-TYPGLNYWRFRDVLVDVIS <b>D</b> SYPRWHAH	284	
Thermus.A4	LQVEILRAHAPGKFVTHNFMG-FFTDLDFAFALAQLDFAS <b>D</b> SYPLGFTD	272	
Thermus.T2	FQVDLLRDNAPGRFITHNFMG-FFTDLDPFALAEDLDFAA <b>D</b> SYPLGFTD	272	
T.neapolitana	LQVEIIRELSGPRFVTHNFMA-GFTDFDHYKISKDLDFAS <b>D</b> NYPLGHTL	273	
B.subtilis	EQAKIIRCYS-DAPITHNGSV-MFS-VDNRMFQNLDFAS <b>D</b> TYAS-QE-	271	
B.circulans	HETKPLKAKNPDLPVTTNLMF-FYEGLYNWKFADILDFL <b>S</b> DSYPTWDA	285	
G.stearothermophilus	TEKEILREVTPDIPVSTNFMG-SFKPLNYFQWAQHVDIVT <b>D</b> SY-----	273	
T.maritima	EEYRAIKKHTPDIPVTTNLIAATFKEWNYFEWAKHMDVA <b>A</b> DNYPGYKED	287	
H.lucentense	LHAALIREANDEWFVTHNFMG-GFS-LDAFRLAADLDFLS <b>D</b> SYPTGFVQ	272	
		. :: : * : : : * . : * . *	
			<b>E313</b>
A.acidocaldarius	E-----TLVPEAVHTAMVHDLNRILK-KPFL <b>L</b> ESTPSVTNW	321	
Thermus.A4	LMPLPP-EEKLRARTGHPDVAFAFHDLRYGVGRGR-FWV <b>M</b> EQQPGPVNW	320	
Thermus.T2	LMPLPQ-EKVQWARTGHPDVAFAFHDLRYGVGRGR-FWV <b>M</b> EQQPGPVNW	320	
T.neapolitana	VFLRAKGESKNPFRVGHDPDIISFSHDLYRGVGRGR-FWV <b>M</b> EQQAGPVNW	322	
B.subtilis	-----NASAFLLNCDLWRNLKQGRPFWIL <b>E</b> TSPSYAAS	304	
B.circulans	-----EEDKLASRIAMMDIVRSIKGGQPFLL <b>E</b> STPSSTNW	323	
G.stearothermophilus	D-----PREGLPIQHMMNDLMRSLRKGQPFIL <b>M</b> EQVTSVNW	311	
T.maritima	-----F-SVISLRHSLIRCLKEGKPFVIL <b>M</b> EQSPSQACW	319	
H.lucentense	DRQPD--PTVDELRAGNPDQVSMNHDLRGAKGKPFWV <b>M</b> EQQPGDINW	319	
		: . : * * : * : *	
			<b>E361</b>
A.acidocaldarius	QAVSKQK-RPGVHVLVSLQAVAHGADSVQYFQWRKSRGS <b>E</b> KFHGA	370	
Thermus.A4	APHNPSP-APGMVRLWTWEALAHGAEVVSYFRWRQAPFA <b>Q</b> EQMHAGLHRP	369	
Thermus.T2	APHNPSP-APGMVRLWTWEALAHGAEVVSYFRWRQAPFA <b>Q</b> EQMQAGFNRP	369	
T.neapolitana	APYNLWP-AEGAVRLWTWQAFAHGAEVVSYFRWRQAPFA <b>Q</b> EQMHSGLLAP	371	
B.subtilis	LESSAYPHADGYLQAEAVSSYALGSQGFYWLWRQQRSG <b>S</b> EISHGSVLSA	354	
B.circulans	QEVSKLK-KPGMHLSSLQAVAHGSDSVQYFQWRKSRGS <b>S</b> EKLHGAVVDH	372	
G.stearothermophilus	RDINVPK-PPGVMRLWSYATIARGADGIMFFQWRQSRAGA <b>E</b> KFHGAMVPH	360	
T.maritima	RWYNPQK-RPGEMRLWSYHALAHGAETLMFFQLRQSKGGV <b>E</b> KFHGAVITH	368	
H.lucentense	PPQSPQP-ADGAMRLWAHHAVAHGADAVVYFRWRRCRQ <b>Q</b> EQYHAGLRRQ	368	
		. * : : * * : : * : . * : . .	

**Figure 2.3:** Particular of Multi-alignment of GH42 sequences. In bold red the invariant carboxylic residues: Glu157, Asp276, Glu313 and Glu361.



The result of the multi-alignment shows several conserved regions and 11 carboxylic residues invariant among all family 42 members. Of these, six Asp (34, 69, 271, 276, 299 and 448) and five Glu (57, 102, 157, 313 and 361) were identified in *A. acidocaldarius*  $\beta$ -galactosidase.

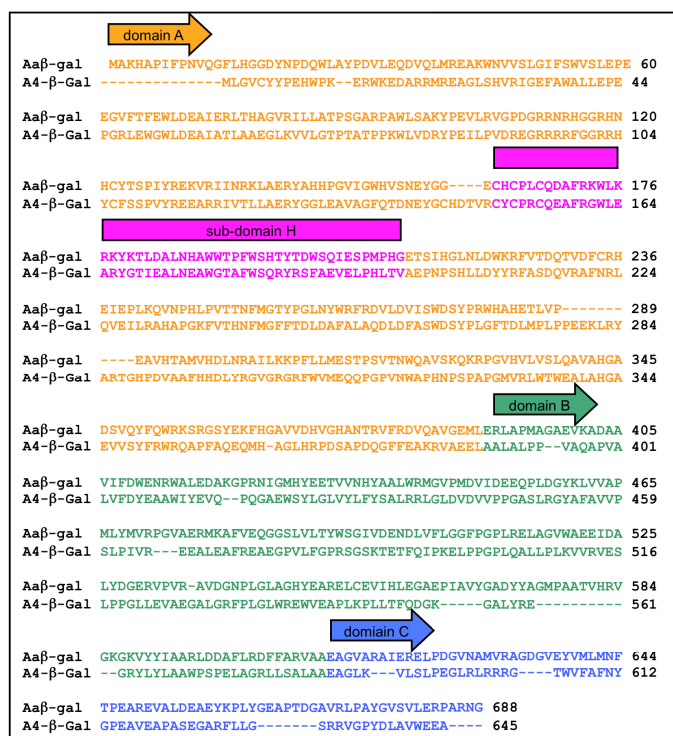
To understand their role in catalysis I prepared a model of the 3D-structure of Aa $\beta$ -gal.

### 3D-structure prediction of Aa $\beta$ -gal

The only 3D structure available in GH42 family is that of A4- $\beta$ -Gal from *T. thermophilus* (Hidaka et al. 2002) showing 32% of identity to Aa $\beta$ -gal.

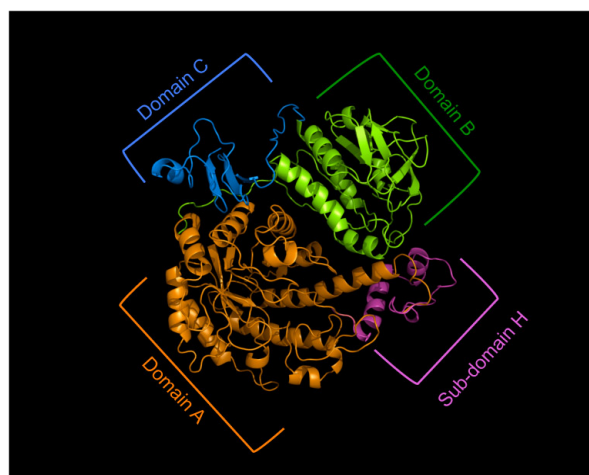
Through 3D-JIGSAW (<http://bmm.cancerresearchuk.org/~3djigsaw/>), an automated software which builds three-dimensional models for proteins based on homologues of known structure, it was possible to obtain the Aa $\beta$ -gal 3D prediction based on A4- $\beta$ -Gal structure. The Aa $\beta$ -gal model (Figure 5.3A) showed high structural similarity to the template with a backbone standard deviation of 0.7 Å.

The 3D-model and the multi-alignment allowed the identification of the highly conserved domains A, B, C and a sub-domain H previously described by Hidaka et al. 2002 (Figure 3.3).



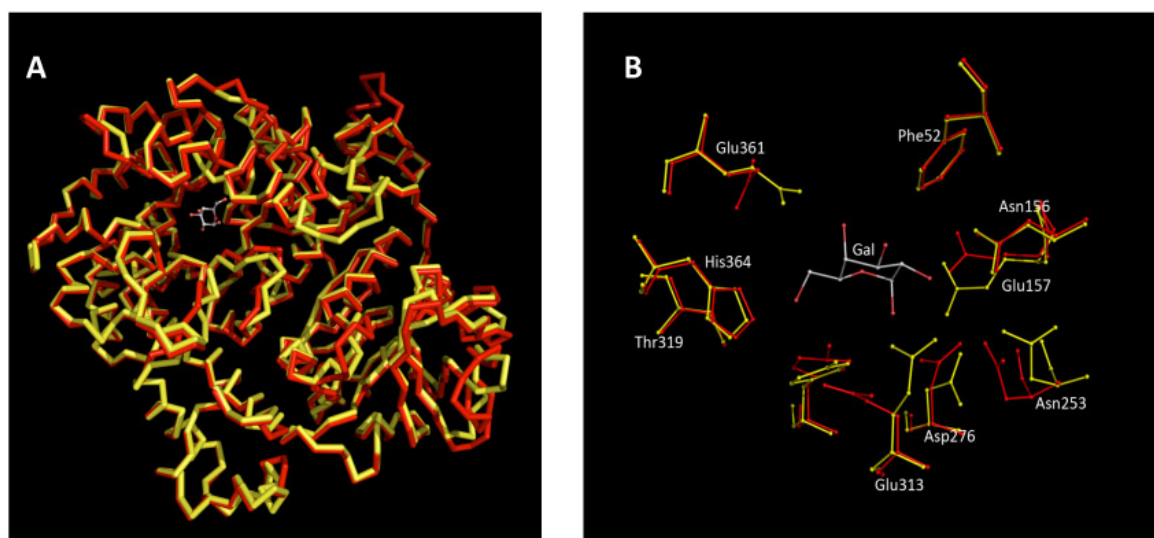
**Figure 3.3:** ClustalW alignment of Aa $\beta$ -gal and A4- $\beta$ -Gal. The domains identified on A4- $\beta$ -Gal template (Hidaka et al. 2002) are indicated with coloured arrows (A,B and C) and a box (sub-domain H).

Domain A shows a  $(\beta/\alpha)_8$  barrel (TIM barrel) supersecondary structure (Figure 4.3), which is also similar to that of a GH-14  $\beta$ -amylase from *Bacillus cereus* (Mikami et al. 1999), containing the active site and an extra region (subdomain H) inserted between  $\beta$ -4 and  $\alpha$ -4 of the barrel. Domain B, showing a  $\alpha/\beta$  fold domain, is involved in the native trimer formation while the function of domain C ( $\beta$  fold domain) is still unknown.



**Figure 4.3:** Aa $\beta$ -gal monomer obtained through 3D-JiGSAW with domains organization and their respective colour code.

The comparison with the 3D-structure of A4- $\beta$ -Gal revealed that Glu 313 and Glu157, corresponded nicely to Glu312 and Glu141, assigned to A4- $\beta$ -Gal as the nucleophile and the acid/base, respectively.



**Figure 5.3:** **A)** Superimposition of backbone trace of Aa $\beta$ -gal (red) and A4- $\beta$ -Gal (yellow). **B)** Detail of A4- $\beta$ -Gal active site (yellow) superimpose with Aa $\beta$ -gal model (red)

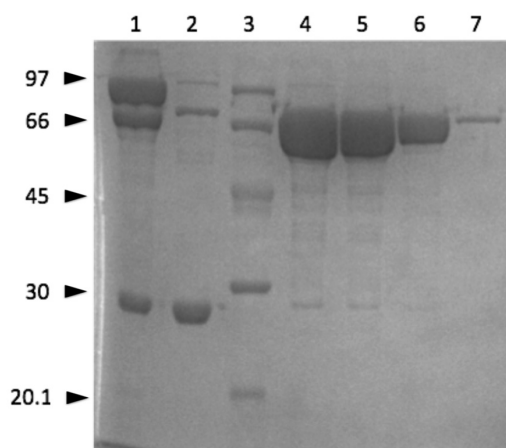
In addition, the search in the model of the highly conserved residues identified by multi-alignment, revealed that only Asp276 and Glu361 are close to the catalytic residues Glu157 and Glu313. The detail of active site in Figure 5.3B, shows the disposition of these residues superimposed to their homologs in *T. thermophilus*.

This analysis suggests that Asp276 and Glu361 could be involved in catalysis in Aa $\beta$ -gal; thus, their function was characterized in detail.

#### Site-directed mutagenesis, expression and purification of Aa $\beta$ -gal mutants

Aa $\beta$ -gal wild type and the mutants Glu157Gly and Glu313Gly were already available in the laboratory cloned in pGEX-2TK plasmid (pGEX-A $\beta$ gal) (Di Lauro et al. 2008).

The mutants Glu313Ser, Glu313Asp, Asp276Gly, Glu361Gly and the double mutant Glu313Gly/Glu361Gly were prepared by site-directed mutagenesis and together with the wild type were expressed and purified exploiting the GST-tag and the thrombin cleavage onto the matrix (Figure 6.3) as described in detail in *Experimental procedures*.



**Figure 6.3:** SDS-PAGE analysis of Aa $\beta$ -gal. Lane 1, Aa $\beta$ -gal bound to the sepharose matrix. Lane 2, sepharose matrix after the thrombin cleavage, Lane 3, molecular weight markers; Lane 4-7, four consecutive elutions of pure Aa $\beta$ -gal (28  $\mu$ g, 22  $\mu$ g, 10  $\mu$ g and 2  $\mu$ g respectively).

The total amount of protein produced is about 35 mg per liter culture for all the mutants except for the Asp276Gly mutant which gave yields of about 10 mg per liter culture; the reasons of this lower expression are still unknown.

After a single-step purification, Aa $\beta$ -gal wild type and mutants were 95% pure by SDS-PAGE and were stored at 4°C.

#### Analysis of steady-state kinetic constants of Aa $\beta$ -gal mutants

The activity of Aa $\beta$ -gal wild type and mutants was analyzed by measuring the steady-state kinetic constants on 2NP- $\beta$ -D-galactopyranoside in 50 mM sodium citrate buffer at pH 5.5 at 65°C (Table 3.3).

	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_{\text{M}}$ (mM)	$k_{\text{cat}}/K_{\text{M}}$ ( $\text{s}^{-1} \text{mM}^{-1}$ )	% of Activity
wt *	$2657 \pm 86$	$5.5 \pm 0.6$	483	100
E313G *	$0.21 \pm 0.04$	$4.8 \pm 0.4$	0.04	0.001
E313D	$2.38 \pm 0.10$	$4.0 \pm 0.6$	0.59	0.008
E313S	ND	ND	ND	-
E157G *	$0.32 \pm 0.02$	$0.11 \pm 0.04$	2.90	0.03
D276G	$54.8 \pm 1.0$	$0.070 \pm 0.005$	2.70	0.03
E361G	$6.0 \pm 0.2$	$5.1 \pm 0.5$	1.18	0.01
E313G/E361G	ND	ND	ND	-

**Table 3.3:** kinetic constants of A $\beta$ -gal wild type and mutants on 2NP- $\beta$ -D-galactopyranoside in 50 mM sodium citrate buffer at pH 5.5 at 65°C. (\*) (Di Lauro et al. 2008). ND: Not detected.

The measured constants showed clearly that all mutations affected dramatically the enzymatic activity; in particular as expected, the removal of Glu313 severely inactivated the enzyme and the  $k_{\text{cat}}/K_{\text{M}}$  for the mutant Glu313Gly was reduced by  $1.2 \times 10^4$ -fold; moreover the mutant Glu313Ser and the double mutant Glu313Gly/Glu361Gly were completely inactive.

The less inactivated mutant is Asp276Gly that showed a loss of activity of 48-fold compared to wild type ( $k_{\text{cat}}$   $54.8 \text{ s}^{-1}$  vs.  $2657 \text{ s}^{-1}$ ) and a  $K_{\text{M}}$  (0.07 mM) 78-fold lower than that of the wild type (5.5 mM).

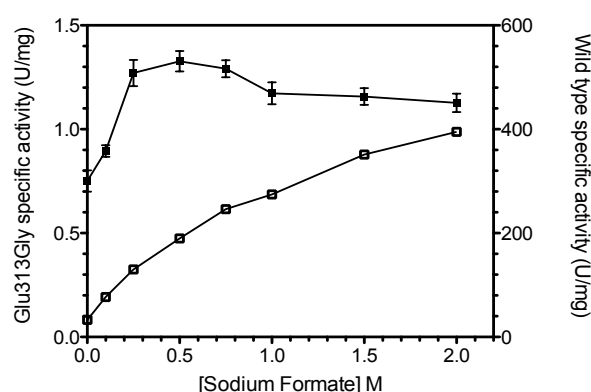
Surprisingly, the kinetic analysis of Glu361Gly showed an affinity constant similar to wild type (5.07 mM vs. 5.5 mM), but a  $k_{\text{cat}}$  reduction of more than 400-fold, suggesting that Glu361 residue could be directly involved in catalysis (Di Lauro et al. 2008).

#### Chemical rescue of A $\beta$ -gal Mutants

As described in the Introduction of this thesis, when the catalytic residues of glycoside hydrolases are replaced by non-nucleophilic amino acids the mutants are severely impaired in their activity. The addition to the reaction mixture of small external anions, such as azide or formate, leads to the reactivation, the so-called *chemical rescue* of the enzymatic activity, in the presence of an activated substrate.

The removal of Glu313 severely inactivated the enzyme (Table 3.3); however, no reactivation of the Glu313Gly mutant was observed when assayed in the presence of concentrations of sodium azide ranging between 20 mM and 1 M. Presumably, the combination of high temperature and sodium azide denatured the mutant protein, precluding the chemical rescue. In contrast, in the presence of sodium formate (up to a final concentration of 2 M), we observed a progressive reactivation of the mutant (Figure 7.3). The  $k_{\text{cat}}$  of the nucleophile mutant Glu313Gly (Table 3.3) increased 21-fold while the specificity constant  $k_{\text{cat}}/K_{\text{M}}$  only 3.5-fold if compared to the value obtained without formate. The latter result is due essentially to the  $K_{\text{M}}$  (30.96 mM) which is about 6.5-fold higher than that obtained in absence of external nucleophile.

Moreover, although the mutant Glu313Asp showed an inactivation lower than that of Glu313Gly ( $k_{\text{cat}}$   $2.38 \text{ s}^{-1}$  vs.  $0.21 \text{ s}^{-1}$ ), the former was not reactivated by any external nucleophile.



**Figure 7.3** Chemical rescue of the Glu313Gly mutant. The wild type (closed symbols) and mutant (open symbols) Aa $\beta$ -gal were assayed on 12 mM 2NP- $\beta$ -D-Gal in 50 mM sodium citrate buffer at pH 5.5 at 65°C in the presence of sodium formate.

This reactivation observed is lower than that of other  $\beta$ -retaining enzymes, which sometimes even approach wild type levels (G Perugino et al. 2003), but it is comparable to that reported for  $\alpha$ -retaining enzymes (Rydberg et al. 2002) (Cobucci-Ponzano et al. 2003) (Chris A Tarling et al. 2003). It was previously described that  $\beta$ -glycosidases mutated in their nucleophile and reactivated by sodium formate are able to promote the synthesis of oligosaccharides and were thereby named glycosynthases (see Introduction) (Moracci et al. 1998). Unfortunately, the Aa $\beta$ -gal mutant Glu313Gly reactivated with this method did not act as a glycosynthase and we could not observe the formation of transgalactosylation products.

The activity of the Glu157Gly mutant was partially rescued in the presence of both sodium formate and sodium azide (0.2 - 2 M) (7- and 3.7- fold increase of  $k_{cat}/K_M$  respectively) when measured in 12 mM 2NP- $\beta$ -D-Gal after 16 h of incubation at standard condition (Table 4.3). In sodium azide, the mutant produced a compound that once isolated and characterized proved to be  $\beta$ -galactosyl azide, indicating that Glu157 is the acid/base of the reaction (Di Lauro et al. 2008). The kinetic analysis in the presence of external nucleophiles (Table 4.3) showed an increase in  $k_{cat}$  of 4.4- and 3.4-fold in formate and azide respectively, while the affinity constant significantly decreased if compared to wild type.

2-Np- $\beta$ -D-gal	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_{\text{M}}$ (mM)	$k_{\text{cat}}/K_{\text{M}}$ ( $\text{s}^{-1} \cdot \text{mM}^{-1}$ )
wt	2657 $\pm$ 86	5.5 $\pm$ 0.6	484
wt + azide 1 M	3424 $\pm$ 230	5.5 $\pm$ 1.1	622
wt + formate 2 M	2489 $\pm$ 230	2.76 $\pm$ 0.68	902
E157G	0.32 $\pm$ 0.02	0.11 $\pm$ 0.04	3
E157G + azide 2 M	1.41 $\pm$ 0.20	0.13 $\pm$ 0.07	11
E157G + formate 2 M	1.10 $\pm$ 0.13	0.05 $\pm$ 0.03	22
E313G	0.21 $\pm$ 0.04	4.8 $\pm$ 0.4	0.04
E313G + azide 2 M	ND	ND	ND
E313G + formate 2 M	4.41 $\pm$ 1.69	30.96 $\pm$ 7.25	0.14
E313D	2.38 $\pm$ 0.10	4.0 $\pm$ 0.6	0.59
E313D + azide 1M	1.02 $\pm$ 0.08	4.26 $\pm$ 1.47	0.23
E313D + formate 2M	1.23 $\pm$ 0.06	3.7 $\pm$ 0.5	0.33

**Table 4.3:** Kinetic constants for the hydrolysis of 2NP- $\beta$ -D-galattopyranoside of Aa $\beta$ -gal wild type and mutants Glu157Gly and Glu313Gly/Asp in 50 mM sodium citrate buffer at pH 5.5 at 65°C. with and without external nucleophiles. ND: not detected.

To determine the role of residues Glu361 and Asp276, the kinetic constants of mutants Glu361Gly Asp276Gly were measured in the presence of sodium azide and sodium formate (Table 5.3).

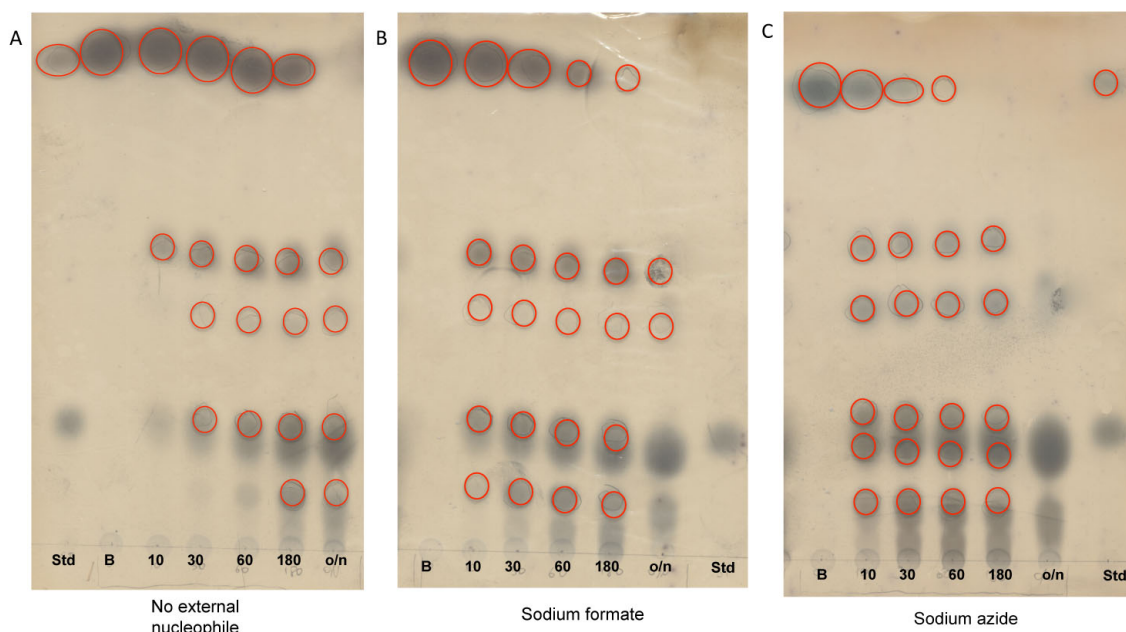
2-Np- $\beta$ -D-gal	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_{\text{M}}$ (mM)	$k_{\text{cat}}/K_{\text{M}}$ ( $\text{s}^{-1} \cdot \text{mM}^{-1}$ )
wt	2657 $\pm$ 86.	5.5 $\pm$ 0.6	484
wt + azide 1 M	3424 $\pm$ 230	5.5 $\pm$ 1.1	622
wt + formate 2 M	2489 $\pm$ 230	2.76 $\pm$ 0.68	902
D276G	54.8 $\pm$ 1.0	0.07 $\pm$ 0.005	783
D276G + azide 2 M	57.83 $\pm$ 0.80	0.12 $\pm$ 0.007	482
D276G + formate 2 M	41.42 $\pm$ 1.69	0.04 $\pm$ 0.02	1035
E361G	6.0 $\pm$ 0.2	5.01 $\pm$ 0.5	1.2
E361G + azide 1 M	167 $\pm$ 17	14.42 $\pm$ 3.01	11.6
E361G + formate 2 M	684.33 $\pm$ 48.98	31.64 $\pm$ 3.57	21.6

**Table 5.3:** Kinetic constants for the hydrolysis of 2NP- $\beta$ -D-galattopyranoside of Aa $\beta$ -gal wild type and mutants Asp276Gly and Glu361Gly in 50 mM sodium citrate buffer at pH 5.5 at 65°C with and without external nucleophiles.

While the  $k_{\text{cat}}$  of the mutant Asp276Gly did not increased significantly in the presence of external nucleophiles, the mutant Glu361Gly was reactivated both by sodium azide and sodium formate showing turnover numbers increased by 29- and 114-fold respectively. Moreover, the use of external nucleophiles with Glu361Gly, although decreased the affinity for the substrate ( $K_{\text{M}}$  14.42 mM and 31.64 mM vs. 5.07 mM), raised the specificity constant by 10- and 18-fold ( $k_{\text{cat}}/K_{\text{M}}$  11.6  $\text{s}^{-1} \text{mM}^{-1}$  and 21.6  $\text{s}^{-1} \text{mM}^{-1}$  vs. 1.2  $\text{s}^{-1} \text{mM}^{-1}$ ). These data suggest that Glu361 could be directly involved in the catalytical machinery of Aa $\beta$ -gal; its role might be better understood after the isolation and characterization of the products of the reaction (work in progress).

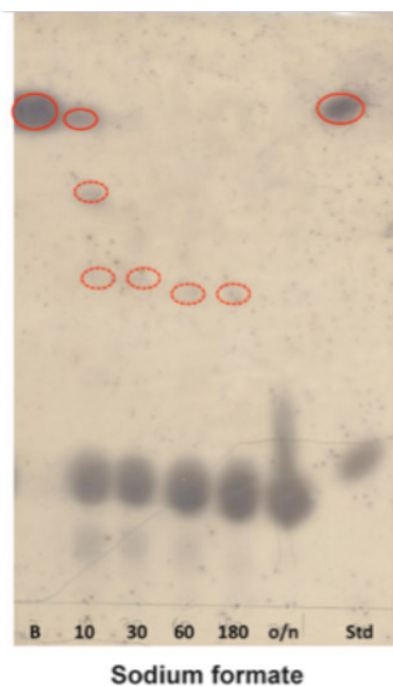
### Transglycosylation activity of Aa $\beta$ -gal mutant Glu361Gly

Reaction mixtures containing Aa $\beta$ -gal Glu361Gly (18  $\mu$ g each), 50 mM sodium citrate buffer (pH 5.5), 80 mM of 2NP- $\beta$ -D-Gal in presence and in absence of external nucleophiles (50 mM sodium azide and sodium formate respectively) were incubated at 65°C. At intervals (from 10 minutes to 16 hours), aliquots were withdrawn from the mixture and analyzed by thin-layer chromatography (TLC) (Figure 8.3A,B,C) as described in “*Experimental procedures*”.



**Figure 8.3:** Time course of the transgalactosylation reaction of Aa $\beta$ -gal Glu361Gly A) no external nucleophiles, B) sodium formate 50 mM, C) sodium azide 50 mM. UV-visible signals are rounded in red. The time scale is in minutes, o/n= 16 hours; std= standards 2-NP- $\beta$ -D-Gal (UV signal) and galactose.

TLCs revealed several spots compatible with the synthesis of transgalactosylation products resulting from autocondensation reactions of 2NP- $\beta$ -D-Gal thereby working as both *donor* and *acceptor*. This result is not surprising because Aa $\beta$ -gal wild type catalyzes transgalactosylation reaction producing a similar products pattern, in sodium formate, but, as reported for a comparison in Figure 9.3, its high hydrolytic activity totally hydrolysed the products in 10 minutes.

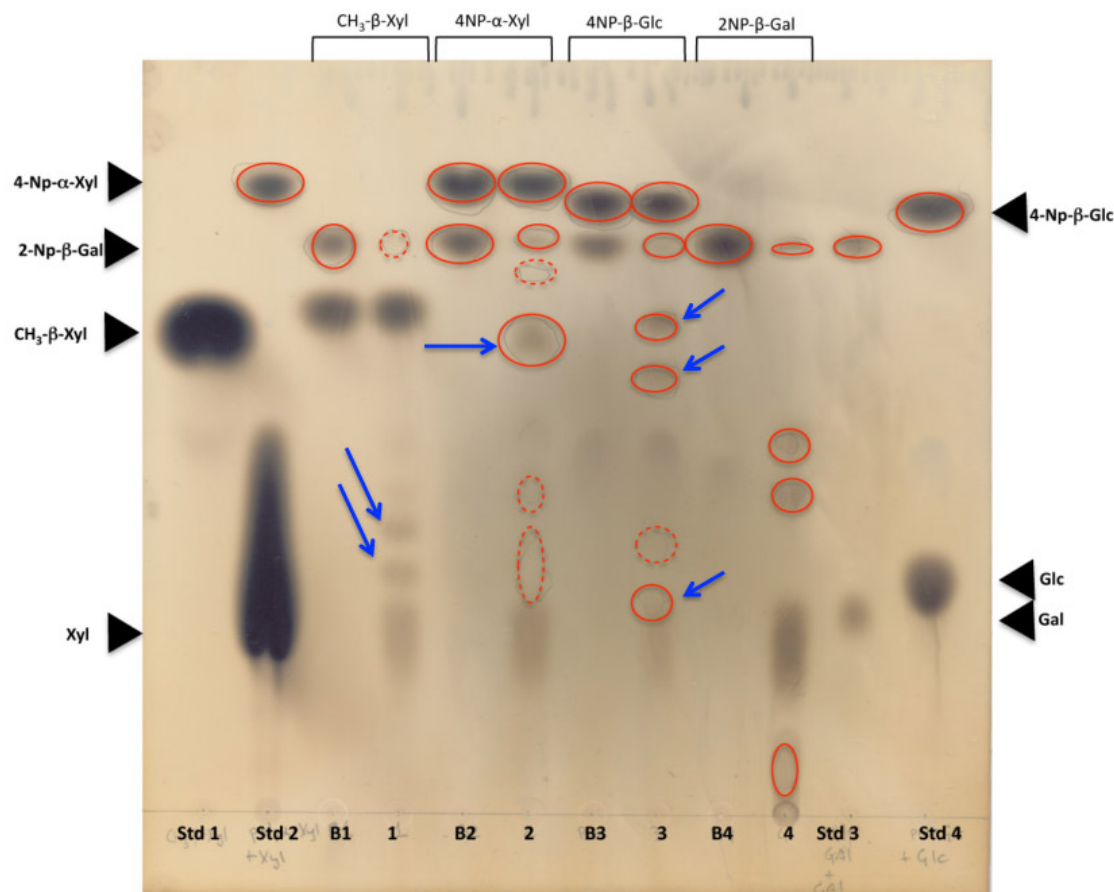


**Figure 9.3:** Time course of the transgalactosylation reaction of Aαβ-gal wild type in sodium formate 50 mM, UV-visible signals are rounded in red. The time scale is in minutes, o/n= 16 hours; std= standards 2-NP-β-D-galactopyranoside (UV signal) and galactose, weak signals are highlighted with dotted line.

Consistently with data of the chemical rescue (Table 5.3) the transglycosylation trials of the mutant Glu361Gly showed a strong reactivation by external nucleophiles with the total conversion of *donor* after 60 minutes in presence of sodium formate and sodium azide respectively. As a comparison, in the absence of external nucleophile, 2NP-β-D-Gal was present up to 180 minutes and the total conversion occurred only after 16 hours.



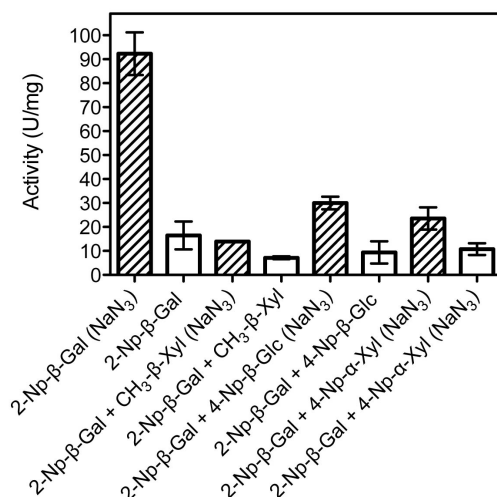
To test the transgalactosylation activity of Glu361Gly on different *acceptors*, reactions were incubated with 2NP- $\beta$ -D-Gal as *donor* and 4NP- $\beta$ -D-glucopyranoside, 4NP- $\alpha$ -D-xylopyranoside and methyl- $\beta$ -D-xylopyranoside, as *acceptors*, in 50 mM sodium citrate buffer (pH 5.5), 50 mM sodium azide at 65° C for 60 minutes (Figure 10.3). The optimal molar ratio between *donor* and *acceptor* for the transgalactosylation (1:2) was established by preliminary tests using different molar ratios (data not shown).



**Figure 10.3:** Transgalactosylation reactions of Glu361Gly with different *acceptors*. 1) 2NP- $\beta$ -Gal 20 mM + CH<sub>3</sub>- $\beta$ -Xyl 40 mM; 2) 2NP- $\beta$ -Gal 20 mM + 4NP- $\alpha$ -Xyl 40 mM; 3) 2NP- $\beta$ -Gal 20 mM + 4NP- $\beta$ -Glc 40 mM 4) 2NP- $\beta$ -Gal 40 mM. UV-visible signals are rounded in red; weak signals are highlighted with dotted line. New products are indicated with a blue arrow.

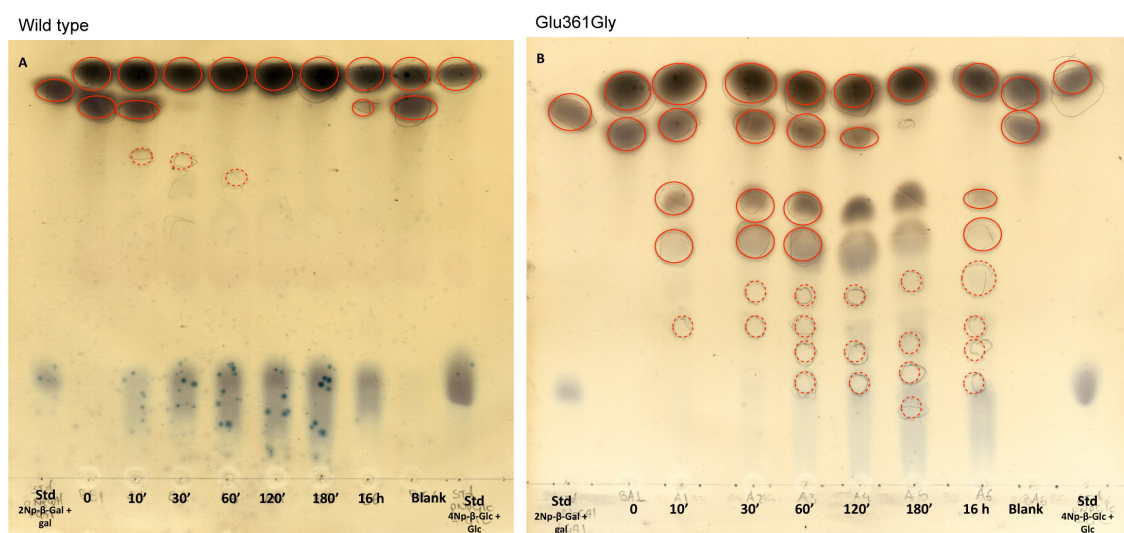
The TLC revealed the synthesis of products different than autocondensation (Figure 10.3, lane 4). In particular, two new products not UV-visible with methyl-xylopyranoside (Figure 10.3, lane 1), a new product UV-visible with 4NP- $\alpha$ -D-xylopyranoside (Figure 10.3, lane 2) and three new products UV-visible with 4NP- $\beta$ -D-glucopyranoside (Figure 10.3, Lane 3), could be observed.

Again, the specific activity of Glu361Gly increased in the presence of 50 mM sodium azide in all the transgalactosylation reaction tested (Figure 11.3).



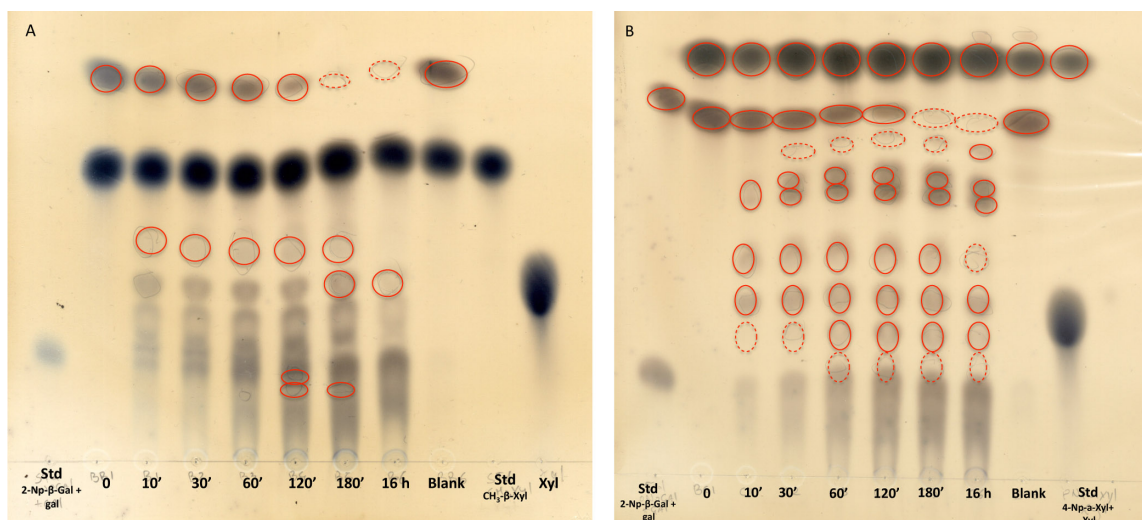
**Figure 11.3:** Activity of mutant Aaβ-gal Glu361Gly with different couples *donor:acceptor*, the reaction were performed in 50 mM sodium citrate buffer (pH 5.5) at 65°C with (hatched histograms) and without 50 mM sodium azide as external nucleophile.

To compare the transgalactosylation abilities of the wild type and mutant Glu361Gly, reactions containing 2NP-β-D-Gal as *donor* and 4NP-β-D-glucopyranoside as *acceptor* and 1.62 Units of both wild type and mutant Glu361Gly were incubated in standard conditions. As previously observed for the autocondensation reaction (See Figure 10.3) the mutant Glu361Gly in the presence of sodium azide had increased transgalactosylation activity if compared to the wild type, with accumulation of products up to 180 minutes (Figure 12.3). In contrast, the wild type catalyzed the total hydrolysis of the substrate after 30 min with no significant synthesis of products.



**Figure 12.3:** Time course of the transgalactosylation reaction of Aaβ-gal wild type (A) and mutant Glu361Gly with sodium azide 50 mM (B). Reactions (Each 1.62 U) in 50 mM citrate buffer with 2-Np-β-D-galactopyranoside and 4-Np-β-D-glucopyranoside at 65°C. UV-visible signals are rounded in red, weak signals are rounded with dotted line.

In addition, the time-course analysis of the reactions performed in the same conditions with other two *acceptors*, methyl- $\beta$ -D-xylopyranoside and 4NP- $\alpha$ -D-xylopyranoside (Figure 13.3), showed the synthesis of several products, different from those obtained by autocondensation. This result further confirms that mutant Aa $\beta$ -gal Glu361Gly catalyzes transgalactosylation reactions with high efficiency.



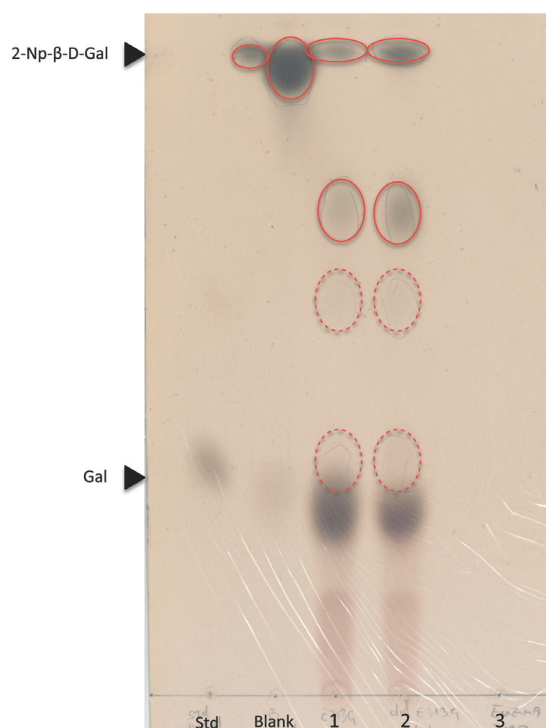
**Figure 13.3:** Time course of the transgalactosylation reaction of Glu361Gly reactions with sodium azide 50 mM, in 50 mM citrate buffer with 2-Np- $\beta$ -D-galactopyranoside and (A) Methyl- $\beta$ -D-xylopyranoside, (B) 4-Np- $\alpha$ -D-xylopyranoside as acceptors at 65°C. UV-visible signals are rounded in red, weak signals are rounded with dotted line.

The sum of these results indicate that the mutant Glu361Gly is a novel glycosynthase able to promote efficiently GOS synthesis. Further studies are required to determine the yields of the glycosynthetic activity of the mutant, however, it is worth noting that this novel glycosynthase was produced by engineering residues not directly involved in catalysis (i.e. the formal catalytic nucleophile). This important result opens remarkable avenues for the further development of the glycosynthase approach allowing to engineer glycosidases whose reaction mechanism involves not only the classical catalytic proposed by Koshland but also other amino acids.

#### Analysis of $\beta$ -galactosynthetic activity of mutant Aa $\beta$ -gal Glu313Gly

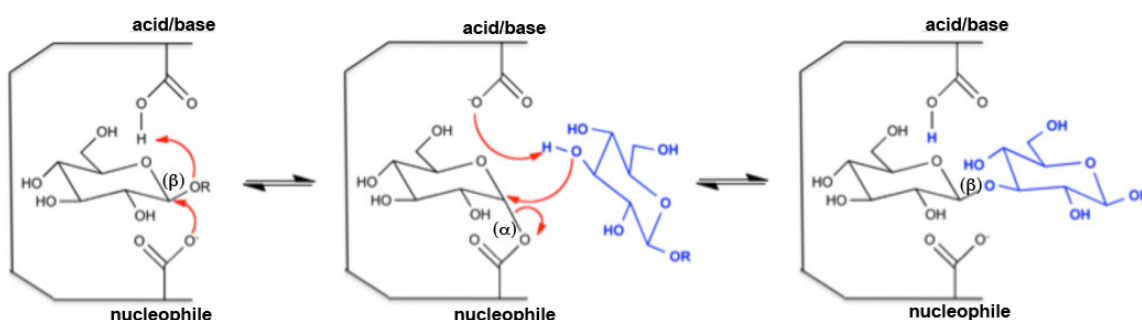
As previously reported, the mutant Glu313Gly incubated with external nucleophiles showed a weak reactivation only in presence of formate 2M (see Table 4.3) and did not act as a glycosynthase in standard conditions (Di Lauro et al. 2008).

However, the analysis by TLC of the reactions performed with very high concentrations (80 mM) of 2NP- $\beta$ -D-Gal (dissolved in 100% DMSO) in 50 mM sodium citrate buffer (pH 5.5), and large amounts (1 mg) of mutant Glu313Gly, surprisingly showed the synthesis of several aryl-derivative products even without the addition of any external nucleophile (Figure 14.3).



**Figure 14.3:** Analysis of Aaβ-gal Glu313Gly transglycosylation reactions (1 mg) in 50 mM citrate buffer with 2-Np-β-D-galactopyranoside at 65°C for 16 h. (1) and (2) are with two different enzyme preparations; (3) the enzyme incubated in the same conditions without substrate. The UV-signal of products are rounded in red, the weak signals are highlighted with dotted line.

The products of reaction were isolated and structurally characterized by Prof. M.M. Corsaro of University of Naples “Federico II”. The  $^1\text{H}$ -NMR analysis showed the presence of a disaccharide 2NP-β-Gal-β(1,3)Gal while other products isolated in small amounts were not characterized.

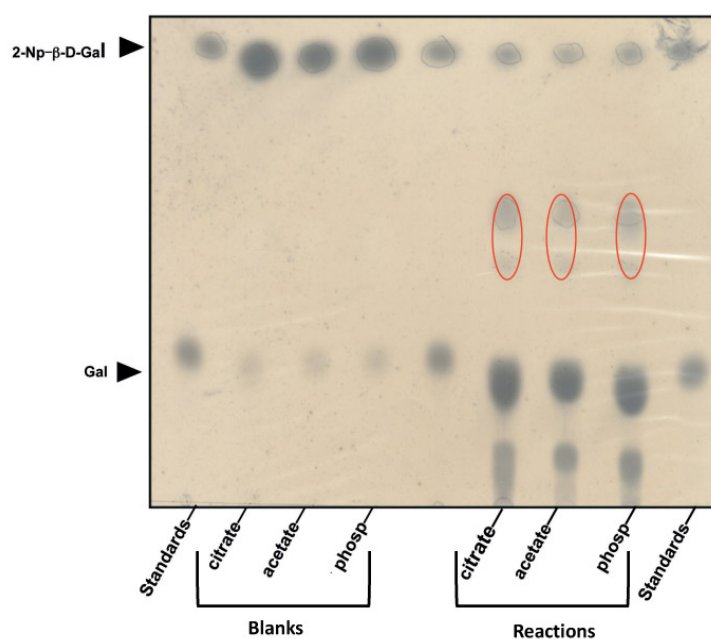


**Figure 15.3:** Transglycosylation reaction of *retaining* glycoside hydrolases

The results obtained, which are typical of transgalactosylation reactions catalyzed by a wild type GHs (Figure 15.3), are totally unexpected. In fact, this mutant, lacking the nucleophile residue and in the absence of any external nucleophile, would not be

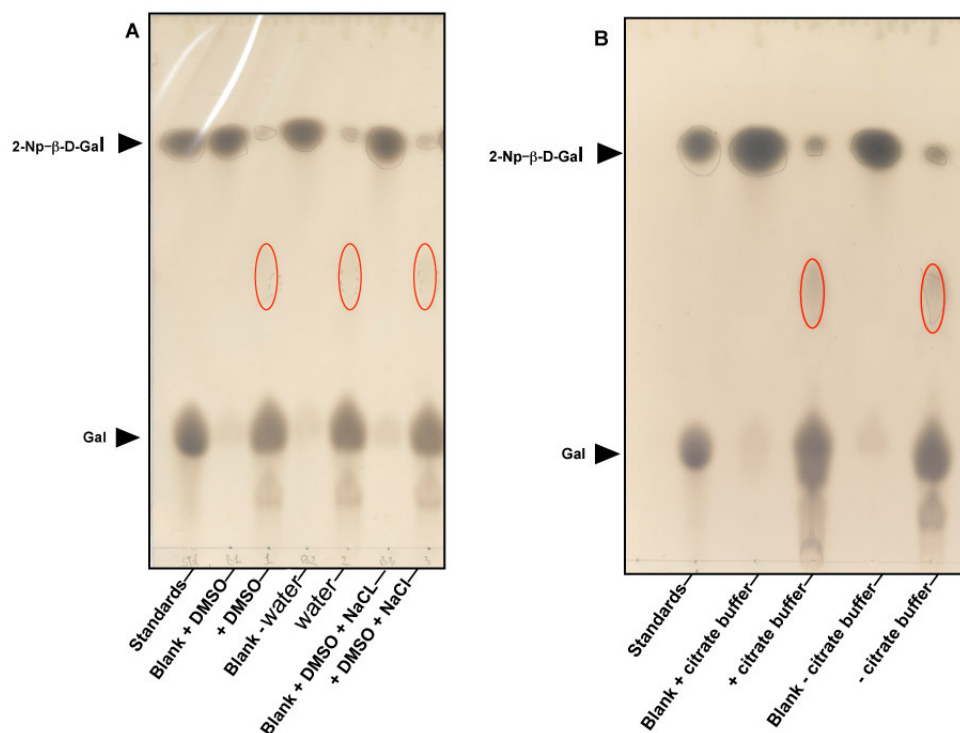
able to form the covalent glycosyl-enzyme intermediate necessary to transfer the glycosyl *donor* to an *acceptor*.

To evaluate if the reactivation observed was due to the citrate buffer, DMSO, or NaCl acting as external nucleophiles, reactions were performed in different conditions, namely sodium phosphate (pH 6.0) and sodium acetate buffer (pH 5.5) (Figure 16.3); DMSO, NaCl, or just in pure water, by removing all this compounds by dialysis (Figure 17.3).



**Figure 16.3** Transglycosylation reactions of Aa $\beta$ -gal Glu313Gly (1 mg) with 80 mM 2-Np- $\beta$ -D-galactopyranoside (solubilised in DMSO at final concentration 600 mM) in 50 mM of different buffer: citrate (pH 5.5), acetate (pH 5.5) and phosphate (pH 6.0). The reactions were incubated at 65°C for 16 h. The two UV-signals are rounded in red.

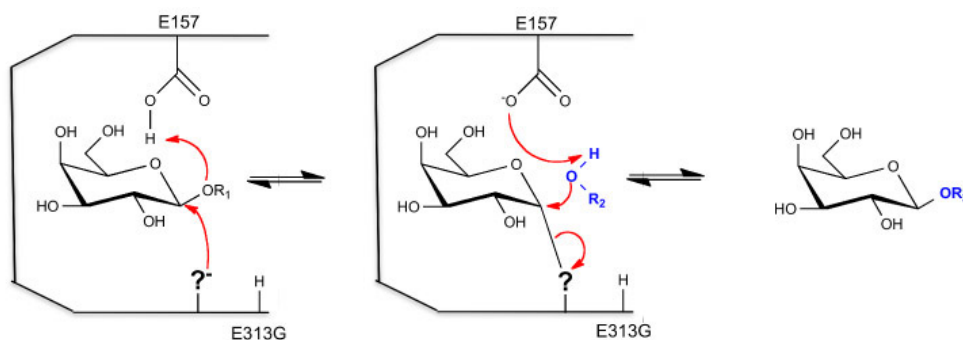




**Figure 17.3:** Reaction mixtures of Aaβ-gal Glu313Gly (A) 0.5 mg of dialysed enzyme with 40 mM 2NP-β-D-Gal, no external buffer; (B) 1 mg of dialysed enzyme with 80 mM 2-Np-β-D-gal in DMSO (final 600 mM). The reactions were incubated at 65°C for 16 h. The two UV-signals are rounded in red.

The analysis of the reactions further confirmed that Gly313Glu is able to catalyze transgalactosylation reactions in the absence of any possible external nucleophile and even in pure water.

A possible explanation is that a not yet identified amino acid in the active site might act as alternative nucleophile when Glu313 is mutated, forming the covalent intermediate  $\alpha$ -galactosyl-enzyme, as showed in Figure 18.3.



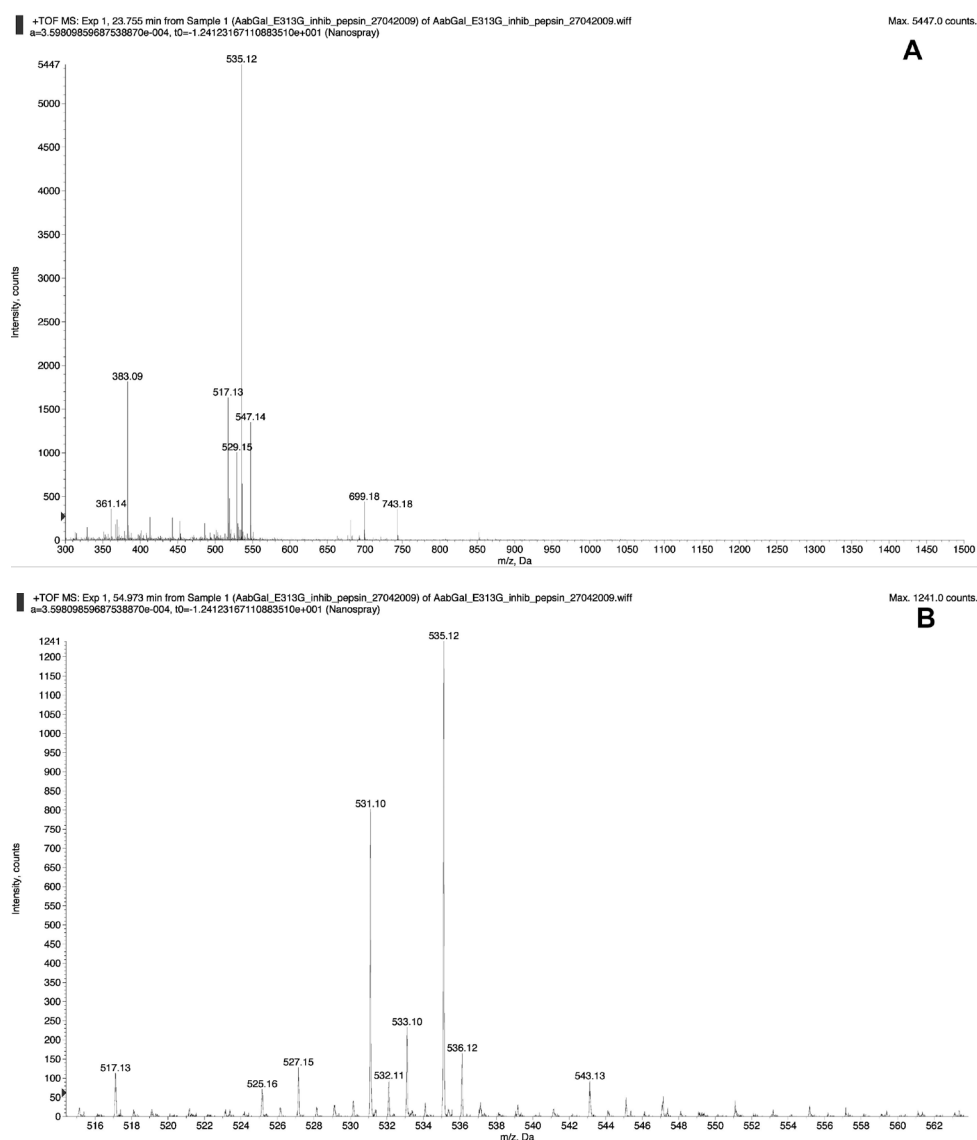
**Figure 18.3:** Mechanism of transgalactosylation proposed in the mutant Glu313Gly;

To test this hypothesis, I followed an approach combining a mechanism-based inhibitor and nano-electrospray ionization tandem mass spectrometry (nano-ESI-

MS/MS), which is usually applied to identify the nucleophile of the reaction (Shaikh et al. 2007).

To identify this putative nucleophile residue, the mutant Glu313Gly was treated with 2,4-dinitrophenyl- $\beta$ -D-2-deoxy-2-fluoro-Gal (2,4DNP-2F-Gal), which was synthesized by Prof. S. G. Withers of the University of British Columbia (Canada).

The Mutant Glu313Gly (0.23 nmol) was incubated with 2,4DNP-2F-Gal in a molar ratio 1:10<sup>3</sup> and 1:10<sup>5</sup> (Enzyme vs. Inhibitor) at 45°C for 2 h as described in “*Experimental procedures*”. The samples were proteolytically digested by pepsin and the resulting peptide mixtures were analyzed by nano-HPLC-ESI-MS/MS in IDA mode. The mass spectrometry analysis did not showed the expected peptide tagged with 2F-Gal at all the conditions tested. However with the molar ratio 1:10<sup>5</sup> a strong signal (M+H<sup>+</sup>) was measured at m/z 535 during the entire analysis (Figure 19.3).



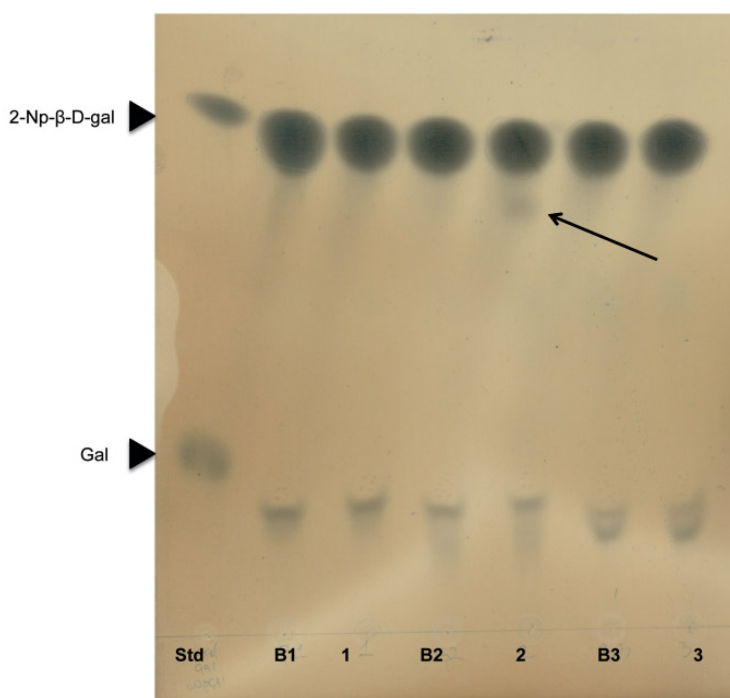
**Figure 19.3:** Nano-ESI mass spectrum of Aa $\beta$ -gal Glu313Gly treated with 2,4DNP-2F-Gal with molar ratio 1:10<sup>5</sup> enzyme vs. inhibitor (A) at 23.75 min. of gradient; (B) at 24.97 min. of gradient

This signal is compatible with the transgalactosylation product 2,4DNP-2F-Gal-(2F-Gal) in sodiate form.

This result is not surprising since it is well known that glycoside hydrolases inactivated by mechanism-based inhibitors are catalytically competent and can transfer the glycosyl moiety from the covalent intermediate to an acceptor. For this reason, mechanism-based inhibitors are often described as being a very slow substrate rather than a true inactivator (Rempel and Withers 2008).

To try to circumvent this problem, in the nucleophile mutant I modified also the acid/base producing the double mutants Glu157Gly/Glu313Gly, Glu157Ser/Glu313Gly and Glu157Gln/Glu313Gly. The aim of this strategy is that of further stabilizing the glycosyl-enzyme intermediate. In fact, the removal of the acid catalyst would not substantially affect the glycosylation step because dinitrophenol, being a very good leaving group, would not require protonic assistance for departure. Instead, the absence of the base catalyst would affect the activation of an acceptor (water or 2,4DNP-2F-Gal itself) severely slowing the hydrolysis of the glycosyl-enzyme intermediate.

As expected, the mutants Glu157Gly/Glu313Gly and Glu157Ser/Glu313Gly were totally inactive in all tested conditions while the activity of the mutant Glu157Gln/Glu313Gly could be chemically rescued, showing the synthesis of a new product in presence of sodium azide as external nucleophile (Figure 20.3, lane 2).

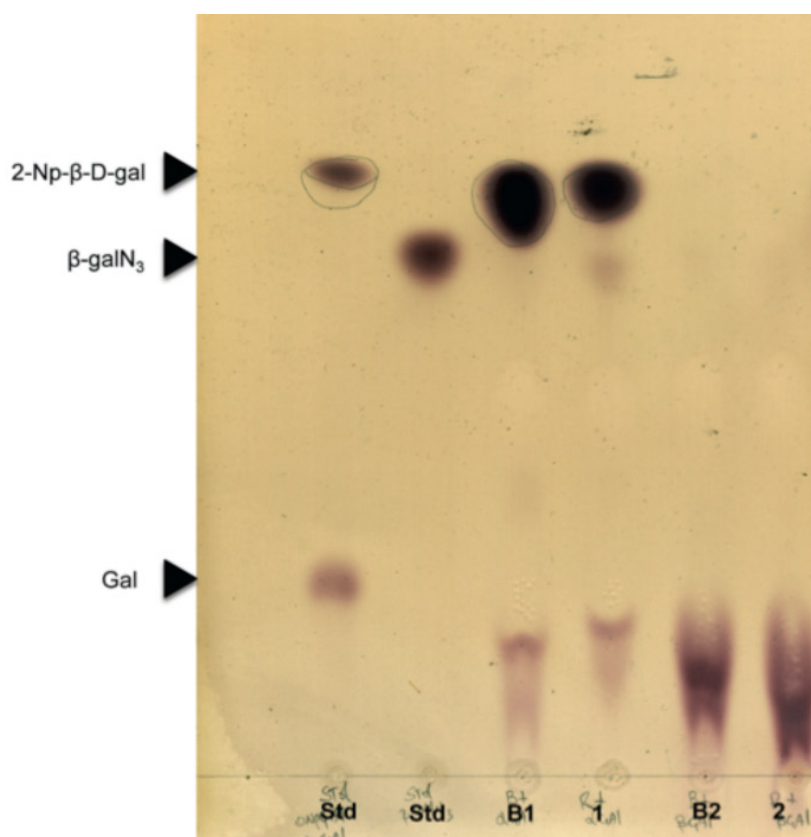


**Figure 20.3:** Chemical rescue of double mutant Aa $\beta$ -gal Glu157Gln/Glu313Gly (1 mg) with 80 mM 2-Np- $\beta$ -D-Gal in 50 mM buffer citrate (pH 5.5) at 65°C for 16 h. 1) without external nucleophile; 2) with 100 mM sodium azide; 3) with 100 mM sodium formate.

To further analyse the new product obtained, aliquots of the reaction were incubated in presence of either the  $\alpha$ -galactosidase from *Thermotoga maritima* (TmGalA) or the



Aa $\beta$ -gal wild type. These enzymes, having 100% stereospecificity for  $\alpha$ - and  $\beta$ -glycosides, respectively, were used to test the anomeric configuration of the products obtained.



**Figure 21.3:** Analysis of the product of the double mutant Aa $\beta$ -gal Glu157Gln/Glu313Gly (1 mg) with 80 mM 2-Np- $\beta$ -D-Gal in 50 mM buffer citrate (pH 5.5) and 100 mM sodium azide at 65°C for 16 h successively incubated with (1) 3  $\mu$ g of Tm $\alpha$ -Gal or (2) with 3  $\mu$ g of Aa $\beta$ -gal at 65°C for 16 h.

The analysis by TLC shows that only the product treated with Aa $\beta$ -gal wild type was hydrolysed indicating a  $\beta$ -glycoside configuration (Figure 21.3). In addition, the product had a migration similar to that of  $\beta$ -galactosyl azide used as standard. The presence of the  $\beta$ -bond indicates that the chemical rescue by sodium azide of Glu157Gln/Glu313Gly occurred during the deglycosylation step (see Figure 4.1B). Therefore, interestingly, these data strongly confirm the formation of a stable covalent intermediate between the 2F-Gal and the unknown novel residue in the active site, indicating that the double mutant Glu157Gln/Glu313Gly is the ideal candidate for the identification of this putative nucleophile.

Preliminary trials incubating the double mutant Glu157Gln/Glu313Gly and 2,4DNP-2F-Gal were performed as above, using a molar ratio 1:10<sup>3</sup> (enzyme:inhibitor), followed by pepsin digestion and analysis by nano-HPLC-ESI-MS/MS.

Unfortunately, only the 86% of protein sequence coverage was obtained by MS/MS sequence data, preventing the identification of the alternative nucleophile. Further

optimization of the digestion conditions is required to improve the sequence coverage.

These results are still preliminary as they did not allowed us yet to define the nature of the putative additional nucleophile. Presumably, this amino acid results from a conformational re-arrangement in the active site of the enzyme as a consequence of the modification of the canonical nucleophile (Glu313). If this hypothesis will be confirmed, it might explain why several GH were recalcitrant to become glycosynthases with the common approaches (see Figure 7.1).

It is worth noting that the mutant Glu313Gly actually acts as glycosynthase producing GOS products, although at high concentrations of both the donor and the catalyst. Therefore, I prepared two mutants, Glu361Gly and Glu3131Gly, working as efficient glycosynthases and we might conclude that Aa $\beta$ -gal was an excellent model system for the production of new catalysts for the synthesis of oligosaccharides.

## **Chapter II: A $\alpha$ -galactosidase from *Thermotoga maritima***

### **Introduction**

## The $\alpha$ -galactosidases

$\alpha$ -D-Galactosidases (EC 3.2.1.22) are exo-acting glycoside hydrolases that cleave  $\alpha$ -linked galactose residues from carbohydrates commonly found in legumes and seeds, such as melibiose, raffinose, stachyose, and gluco- or galacto- mannans (Meier et al. 1982).

These enzymes are classified on the basis of their substrate specificity: group I  $\alpha$ -galactosidases hydrolyze oligosaccharides, such as melibiose, stachyose, raffinose, and verbacose; group II  $\alpha$ -galactosidases are active on polysaccharide substrates, such as galactomannan and glucomannan. In the CAZy classification,  $\alpha$ -galactosidases are found in families GH4, GH27, GH36, and GH57. The enzymes from eukaryotes are predominantly grouped into family GH27, whereas those from microbial sources are primarily grouped into families GH4, GH36, and GH57. Of these, families GH27 and GH36 are thought to share a common ancestral gene, forming the glycoside hydrolase clan GH-D with a common TIM barrel structure (Comfort et al. 2007).

### Applications of $\alpha$ -galactosidases

$\alpha$ -Galactosidases are used in several biotechnological applications, including the pretreatment of animal feed to hydrolyze non-metabolizable sugars, thereby increasing the nutritive value, in the degradation of raffinose to improve the crystallization of sucrose, in the processing of soy molasses and soybean milk, in the improvement of the viscosity and gelling properties of galactomannan for the stimulation of oil/gas wells through hydrolysis of the propanant matrix (Ghazi et al. 2003; Fridjonsson et al. 2001; Tayal et al. 1999; McCutchen et al. 2000). In medical applications they convert B-type blood antigens to produce type O blood (Hata et al. 2004) and are used in the treatment of Fabry's disease, an X-linked lysosomal storage disorder (LSD) caused by a deficiency in  $\alpha$ -galactosidase A. This deficiency causes systemic accumulation of galactosylsphingolipid moieties, especially globotriaosylceramide (Gb3) (Yoshimitsu et al. 2004). Here, the enzyme replacement therapy has been shown to successfully treat the disease (Eng et al. 2001).

Oligosaccharides containing terminal  $\alpha$ -Gal-(1-3)- $\beta$ -Gal sequence ( $\alpha$ -Gal epitope), synthesized *in vivo* by a  $\alpha$ -1,3-galactosyltransferase ( $\alpha$ -1,3GT), have been identified as the major xenoactive antigen responsible for hyperacute rejection (Galili 2001). This immune reaction, mediated by the natural anti-Gal antibody that specifically binds to the  $\alpha$ -Gal epitope, represents the most serious barrier to the transplantation of organs from nonprimate to humans. Since anti-Gal is produced in large amounts in all humans (unless they are severely immunocompromized)  $\alpha$ -Gal epitope has potential for a variety of clinical uses, particularly in the areas of viral and cancer vaccines. Indeed, synthetic  $\alpha$ -Gal epitopes may be coupled to the vaccines that lacks these types of carbohydrate chains (Macher et al. 2008).

Therefore, methods allowing the facile and efficient synthesis of  $\alpha$ -galactooligosaccharides are urgently needed. As previously mentioned (see General Introduction) glycosynthases (GS), represent a reliable alternative to the chemical synthesis of carbohydrates (Mackenzie et al. 1998); (Moracci et al. 1998). Recently it

was reported that two phylogenetically unrelated  $\alpha$ -L-fucosidases could be converted in efficient  $\alpha$ -fucosynthases by reactivating the nucleophile mutants with  $\beta$ -fucosyl-azide (Figure 8.1), as an alternative to the classical  $\beta$ -fluoride derivative donor (Cobucci-Ponzano et al. 2009).

To further confirm that the use of  $\beta$ -azide derivatives could be of general applicability for  $\alpha$ -glycosynthases,  $\beta$ -galactosyl-azides were considered as possible donors for  $\alpha$ -galactosynthases,. To this aim, the  $\alpha$ -galactosidase (TM1192) from the hyperthermophilic bacterium *Thermotoga maritima* (TmGalA) was chosen as model system (Cobucci-Ponzano et al. 2010).

TmGalA is a thermophilic and thermostable enzyme of 552 residues with a molecular weight of 64 kDa active at 90°C and an half-life of 6.5 h at 85°C (Liebl et al. 1998). It belongs to family GH36 and shows high substrate specificity towards galactosides. Comfort and co-workers in 2007 characterized the recombinant form of TmGalA demonstrating its *retaining* reaction mechanism and identifying its catalytic residues Asp327 and Asp387 as the nucleophile and the acid/base, respectively (Comfort et al. 2007). Moreover, the three-dimensional structure at 2.3 Å resolution has also been deposited (PDB ID: 1ZY9). All the information available make TmGalA as a interesting target for the development of a new glycosynthase.

## **Chapter II**

### **Results and Discussion**

### Kinetic characterization and chemical rescue of TmGalA nucleophile mutants.

The functional role of catalytic nucleophile in the  $\alpha$ -galactosidase from *T. maritima* (TmGalA) has been assigned to Asp327, based on structural alignment with other GH36 enzymes, analysis of the activity vs pH, azide rescue of the enzymatic activity, and characterization of the products obtained with the mutant Asp327Gly. These data prompted us to test if mutants in Asp327 could act as  $\alpha$ -galactosynthases. To this aim, we analyzed the mutants Asp327Ala and Gly, previously described (Comfort et al. 2007), and the newly prepared mutant Asp327Ser. The Asp327Ala and Ser mutants expressed in *E. coli* were preliminary assayed in permeabilized cells at 65°C in 50 mM sodium acetate, pH 5.0, and resulted completely inactive as expected. However, no  $\alpha$ -galactosidase activity could be rescued by adding 0.1-2.0 M sodium azide suggesting that the two mutations had a detrimental effect on the activity and/or the stability of TmGalA. Therefore, these mutants were not characterized any further.

It was previously reported that Asp327Gly mutant had its hydrolysis rate and catalytic efficiency reduced by 200-300-fold on 4-nitrophenyl- $\alpha$ -D-galactopyranoside (4NP- $\alpha$ -D-Gal) at 37°C if compared to wild type TmGalA (Comfort et al. 2007). This inactivation is lower than that commonly observed in other glycoside hydrolases mutated in the nucleophile (Viladot et al. 1998; Hrmova et al. 2002; Bravman et al. 2003), but it is not uncommon among  $\alpha$ -glycosidases (Knegt et al. 1995; Shallom et al. 2002; Tarling et al. 2003). However, when steady state kinetic constants were measured at 65°C, we observed a larger inactivation:  $10^3$ - and  $2.6 \times 10^3$ -fold of the  $k_{cat}$  and  $k_{cat}/K_M$ , respectively, if compared to the wild type (Table 1.2). Possibly, the inactivation produced by the mutation Asp327Gly was underestimated at 37°C because of the lower specific activity of the wild type at this temperature.

	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ (mM)	$k_{cat}/K_M$ (s <sup>-1</sup> mM <sup>-1</sup> )
<b>Wild type</b>	65±2	0.04±0.02	1585
<b>D327G</b>	0.060±0.003	0.10±0.04	0.6
<b>D327G<sup>1</sup></b>	0.060±0.003	0.10±0.04	0.6
<b>D327G in 500 mM NaN<sub>3</sub></b>	2.07±0.04	0.12±0.01	17

**Table 1.4:** Steady-state kinetic constants of wild type and mutant TmGalA. Assays were performed in 50 mM sodium acetate, pH 5.0 at 65°C on 2.5 mM 4NP- $\alpha$ -D-Gal; <sup>1</sup>Assays were performed at the same conditions but in 50 mM sodium citrate-phosphate, pH 5.0

Similar results were obtained in both sodium acetate and sodium citrate-phosphate buffers, indicating that these ions did not rescue the hydrolytic activity of the mutant and that this is the basal activity of Asp327Gly (Table 1.4).

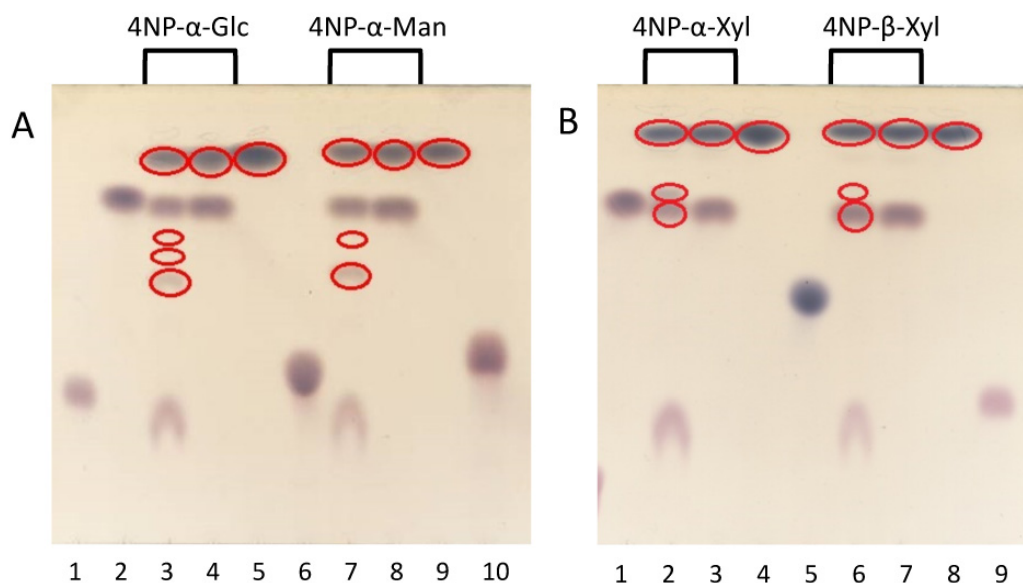
The chemical rescue of the activity of the Asp327Gly in the presence of sodium azide was similar to that previously reported at 37°C (about 30-fold) (Table 1.4.), and the mutant and wild type TmGalA showed similar dependence on temperature of the  $\alpha$ -galactosidase activity (Comfort et al. 2007).

## Characterization of the $\alpha$ -galactosynthet activity of mutant TmGalA Asp327Gly

### $\alpha$ -galacto-oligosaccharide synthesis by the mutant Asp327Gly.

By following the same approach reported for two  $\alpha$ -fucosynthases (Cobucci-Ponzano et al. 2009) TmGalA Asp327Gly mutant was tested for promoting transgalactosylation reactions from  $\beta$ -galactosyl-azide ( $\beta$ -GalN<sub>3</sub>), which was promptly chemically synthesized from galactose by Dr. E. Bedini of the University of Naples "Federico II" in three steps and 85% overall yield (see *Experimental procedures*).

Incubations of the mutant at 65°C in 50 mM sodium acetate buffer, pH 5.0, in the presence of  $\beta$ -GalN<sub>3</sub> as unique substrate (at concentrations 1-14 mM) did not lead to any product by inspection of runs on TLC, indicating that no autocondensation reactions were catalyzed at these conditions (not shown). Therefore, different glycosides were included in the reaction as possible acceptors at 1:1 and 1:10 donor: acceptor molar ratios. Most of the acceptors used did not lead to transgalactosylation products at these conditions, suggesting that Asp327Gly is rather selective in the -1 sub-site of the catalytic center, according to current nomenclature (Davies et al. 1997). However, UV-visible products showing a different migration compared to substrates were clearly observed by TLC, when we used 4NP- $\alpha$ -Glc, 4NP- $\alpha$ -Man, 4NP- $\alpha$ -Xyl, and 4NP- $\beta$ -Xyl as acceptors at a 1:1 molar ratio with the donor (Figure 1.4).

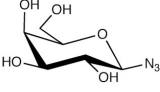
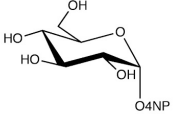
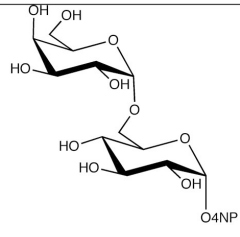
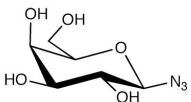
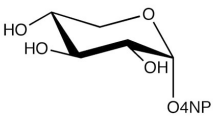
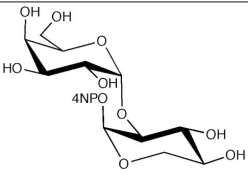
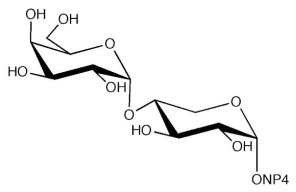


**Figure 1.4:** Transgalactosylation reactions of the Asp327Gly mutant (20  $\mu$ g) in 50 mM sodium acetate buffer, pH 5.0, at 65°C for 16 using 14 mM *donor* and 14 mM *acceptor* (molar ratio 1:1) – (A) Reactions with 4-Np- $\alpha$ -Glc and 4-Np- $\alpha$ -Man acceptors. Lane 1: Gal standard ; lane 2:  $\beta$ GalN<sub>3</sub> standard; lane 3: reaction  $\beta$ GalN<sub>3</sub> , 4-Np- $\alpha$ -Glc hrs; lane 4: blank control ; lane 5: standard 4-Np- $\alpha$ -Glc ; lane 6: Glc standard ; lane 7: reaction containing  $\beta$ GalN<sub>3</sub> , 4-Np- $\alpha$ -Man, lane 8: blank control; lane 9: standard 4-Np- $\alpha$ -Man; lane 10: Man standard. (B) Reactions with 4-Np- $\alpha$ -Xyl and 4-Np- $\beta$ -Xyl acceptors. Lane 1:  $\beta$ GalN<sub>3</sub> standard; lane 2: reaction containing  $\beta$ GalN<sub>3</sub> , 4-Np- $\alpha$ -Xyl; lane 3: blank control; lane 4: standard 4-Np- $\alpha$ -Xyl; lane 5: Xyl standard ; lane 6: reaction containing  $\beta$ GalN<sub>3</sub> , 4-Np- $\beta$ -Xyl, ; lane 7: blank control ; lane 8: standard 4-Np- $\beta$ -Xyl ; lane 9: Gal standard. UV-visible reaction products and substrates are highlighted with red ovals.

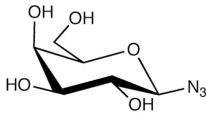
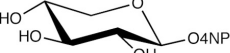
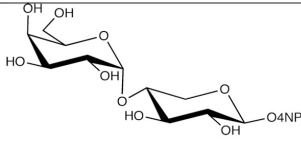
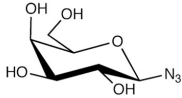
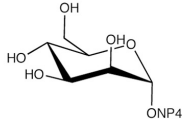
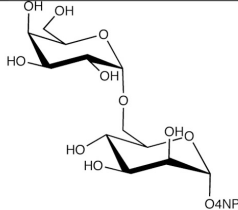
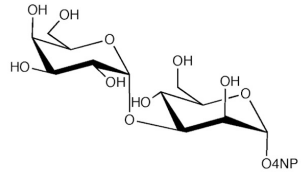


No transgalactosylation products could be observed when 4NP- $\alpha$ -D-Xyl and 4NP- $\beta$ -D-Xyl were used at a 1:10 donor: acceptor molar ratios; possibly, at these conditions, they compete with  $\beta$ -GalN<sub>3</sub> in the +1 donor binding site.

The transgalactosylation efficiency was measured by analysing aliquots of the reaction by HPAEC-PAD. The different products were isolated and characterized by Prof. M.M. Corsaro of the University of Naples "Federico II" and the results are summarized in Table 2.4 and 3.4. In the presence of 4NP- $\alpha$ -Glc, only compound **1** ( $\alpha$ -Gal-(1-6)- $\alpha$ -Glc-4NP) was synthesized in 33% yield (Table 2.4, reaction I). Higher yields were obtained when 4NP- $\alpha$ -Xyl and 4NP- $\beta$ -Xyl were used as acceptors (40% and 38% for reactions II and III, respectively), obtaining, with the former, compounds **2** and **3** ( $\alpha$ -Gal-(1-2)- $\alpha$ -Xyl-4NP and  $\alpha$ -Gal-(1-4)- $\alpha$ -Xyl-4NP, respectively) and only compound **4** ( $\alpha$ -Gal-(1-4)- $\beta$ -Xyl-4NP) with 4NP- $\beta$ -Xyl acceptor. It is worth mentioning that the Asp327Gly mutant showed good regioselectivity with transfer of the galactose moiety exclusively onto a single OH. When 4NP- $\alpha$ -Glc was the acceptor, the enzyme transgalactosylated exclusively the primary alcohol at the C6; the other compounds observed by TLC were present in negligible amounts and could not be isolated. The regioselectivity on xyloside acceptors differs on the basis of the anomeric configuration with main products of transgalactosylation on the OH at the C2 and C4 groups of 4NP- $\alpha$ -Xyl (Table 2.4 reaction II) and 4NP- $\beta$ -Xyl (Table 3.4, reaction III), respectively. Possibly, this is the result of the different binding of these molecules on the -1 sub-site. In reaction IV, with 4NP- $\alpha$ -Man as acceptor, we could isolate two products, **5** and **6**, corresponding to  $\alpha$ -Gal-(1-6)- $\alpha$ -Man-4NP and  $\alpha$ -Gal-(1-3)- $\alpha$ -Man-4NP, respectively, with total yields of 51%. Again, the primary alcohol at the C6 was the preferred functional group, while the transgalactosylation product on the OH at the C3 was synthesized at very low amounts as shown by the relative molar ratio of about 9:1 (Table 3.4, reaction IV).

Synthetic products of TmGalAD327G				
Reaction	Donor	Acceptor	Products	Relative ratios (%)
I				100
			<p><b>1</b></p> <p>Total transgalactosylation efficiency 33%</p>	
II				73
				27
			<p><b>3</b></p> <p>Total transgalactosylation efficiency 40%</p>	

**Table 3.4:** Summary table of the synthetic reactions of TmGalA D327G in presence of  $\beta$ -GalN3 as donor and: I) 4Np- $\alpha$ -Glc; II) 4Np- $\alpha$ -Xyl as acceptor.

III	 	 <p><b>4</b></p>	100
Total transgalactosylation efficiency 38%			
IV	 	 <p><b>5</b></p>	92
 <p><b>6</b></p>	8	Total transgalactosylation efficiency 51%	

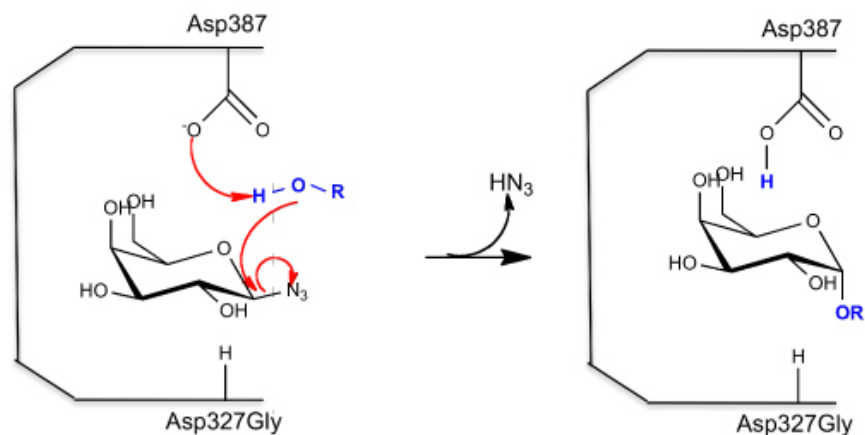
**Table 3.4:** Summary table of the synthetic reactions of TmGalA D327G in presence of  $\beta$ -GalN3 as donor and: I) 4Np- $\beta$ -Xyl; IV) 4Np- $\alpha$ -Man.

### Reaction mechanism of $\alpha$ -D-galactosynthase

We demonstrated that the TmGalA mutant Asp327Gly is a novel  $\alpha$ -galactosynthase, the first produced so far, to the best of our knowledge, and the third example of  $\alpha$ -glycosynthases along with  $\alpha$ -glucosynthase and  $\alpha$ -fucosynthase (Okuyama et al. 2002; Cobucci-Ponzano et al. 2009; Wada et al. 2008). The approach reported here, exploiting  $\beta$ -glycosyl-azides as donors, is the same recently proposed for two  $\alpha$ -fucosynthases from family GH29 (Cobucci-Ponzano et al. 2009).

The mechanism followed by Asp327Gly  $\alpha$ -glycosynthase is shown in Figure 2.4: the cavity created by the mutation that removed the side chain of Asp327 allowed the access to the active site of the  $\beta$ -Gal-N<sub>3</sub>. Then, the galactose moiety is transferred to

the acceptor activated by the base catalysis of the Asp387 residue; the disaccharide product, showing the newly formed  $\alpha$ -bond, cannot be hydrolyzed by the mutant and thus accumulates in the reaction mixture.



**Figure 2.4:** Reaction mechanism of the  $\alpha$ -galactosynthase TmGalA Asp327Gly

This approach follows the same principle proposed for the first time by Withers and collaborators for  $\beta$ -glycosynthases exploiting fluoro-glycoside derivatives with an anomeric configuration opposite to that of the 'natural' substrate of the enzyme (Mackenzie et al. 1998). Here, following from GH29  $\alpha$ -fucosynthases (Cobucci-Ponzano et al. 2009), we confirm that  $\beta$ -azide-glycoside derivatives can be usefully exploited also by GH36  $\alpha$ -galactosynthases, demonstrating that azide derivatives are a valid alternative to fluoro derivatives for the production of  $\alpha$ -glycosynthases. In particular,  $\beta$ -Fuc- $N_3$  and  $\beta$ -Gal- $N_3$  are more stable than their  $\beta$ -Fuc-F and  $\beta$ -Gal-F counterparts, because sodium azide is a less effective leaving group than fluoride. Therefore,  $\beta$ -glycosyl-azides are not degraded spontaneously at the operational conditions and, nonetheless, are sufficiently reactive donors. We demonstrated this property with the observation that  $\beta$ -fucosyl-,  $\beta$ -galactosyl-, and  $\beta$ -mannosyl-fluorides, being 6-deoxyhexopyranosides and/or showing axial substituents on the C2 and C4, are easily activated than hexopyranosides with C4 equatorial substituents, respectively (Overend 1972; Albert et al. 2000).

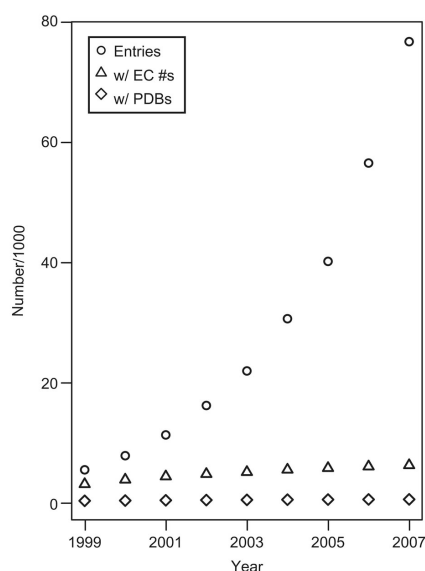
It is likely that, the instability of  $\beta$ -fluoride derivatives have hampered the production of  $\alpha$ -glycosynthases; as an attractive alternative,  $\beta$ -azide derivatives can serve as substrates for the production of novel  $\alpha$ -glycosynthases from other unrelated families of glycosidases.

## **Chapter III: A $\beta$ -glycosidase from *Sulfolobus solfataricus***

### **Introduction**

## A new $\beta$ -glycosidase from *S. solfataricus*

As previously mentioned (see General Introduction), the number of CAZy families is continuously increasing. This was achieved thanks to the advent of new DNA sequencing techniques and subsequent sophisticated computer-aided sequence annotation procedures combined with new biochemical characterization. Interestingly, although the number of CAZy sequences increased 14-fold in the last 8 years, the number of enzymatic and structural characterization only doubled in the same time span, with at present <10% of total proteins in the CAZy database that have been characterized enzymatically (Figure 1.5)(Cantarel et al. 2009). This contrast clearly shows that, in comparison with highly automated sequencing techniques, enzymatic characterization of novel CAZymes is a longer and laborious process representing the limiting step for the full exploitation of genome sequencing efforts. Therefore, the characterization of novel carbohydrate active enzymes is of extremely valuable to increase our knowledge on this topic.



**Figure 1.5** The number of protein containing CAZy modules were noted in December of the years 1999–2007. Within this set (Open circle), the number of enzymatically characterized proteins (triangle) and those with solved structures (open diamond) were also counted (Cantarel et al. 2009)

In the laboratory in which I have performed my thesis, a novel GH from the hyperthermophilic Archaeon *S. solfataricus* was cloned, expressed and characterized enzymatically in detail. This enzyme, encoded by the ORF SSO1353 displays sequence similarity to several unknown proteins from the three domains of life (Archaea, Bacteria, and Eukarya) including the human non-lysosomal glucosylceramidase, also known as  $\beta$ -glucosidase 2 (GBA2) (Boot et al. 2007), an enzyme involved in an alternative catabolic pathway of glucosylceramide (van Weely et al. 1993).

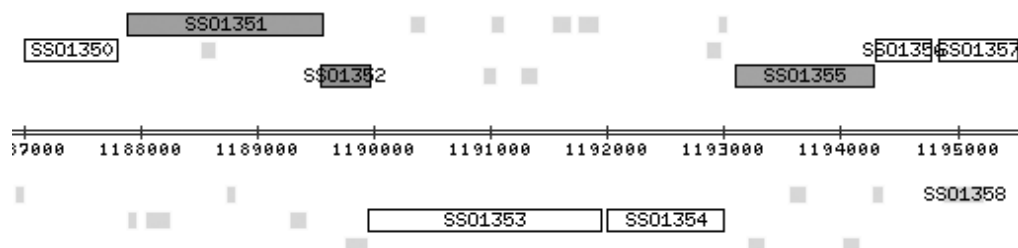
Today there is no 3-D structure representative for the homologs of SSO1353 and only two sequences are characterized: glucosylceramidase 2 (GBA2) from *H. sapiens* and *M. musculus* respectively.

## **Chapter III**

### **Results**

## Isolation of ORF SSO1353

The inspection of the genomic sequence of the archaeon *S. solfataricus*, strain P2, revealed an ORF downstream of the gene *sso1354* encoding for an endoglucanase (Figure 2.5) (Maurelli et al. 2008).



**Figure 2.5:** ORFs scheme of the region near to SSO1353 in the *S. solfataricus* genome

Sso1353 is presently annotated as an hypothetical protein while the other ORFs in this cluster, *sso1351*, *sso1352*, and *sso1355*, are a putative permease, a transcriptional regulator, and a carboxypeptidase, respectively. ORFs *sso1354* and *sso1353* are transcribed in the same direction and are separated by 57 bp in which the latter ORF is preceded by a putative promoter formed by an AT-rich box A (centered at -30 nt from ATG) and a TFB-responsive element (centered at -38 nt) (not shown). Northern blot analysis showed that *sso1353* is expressed as an isolated gene (not shown) and the absence of a clear Shine-Dalgarno-like motif in the intergenic region suggests that *sso1353* gene is translated as a leaderless gene (Torarinsson et al. 2005). Initial gapped BLAST searches (Altschul et al. 1997) revealed that SSO1353 is similar to proteins of unknown function from Archaea, Bacteria, and Eukarya and, to a lesser extent, to eukaryotic non-lysosomal bile acid  $\beta$ -glucosidases. The higher sequence identity scores (>31%) were with archaeal proteins with the highest (86%) with loci *sso1948* from *S. solfataricus*, strain P2, and M1425\_0924 and M1627\_099 from *S. islandicus*, strains M.14.25 and L.S.2.15, respectively. Interestingly, these highly similar genes also lie downstream to a locus encoding for an endoglucanase (SSO1949 in *S. solfataricus* (Huang et al. 2005)), suggesting that gene duplication occurred in these organisms. Among non-lysosomal bile acid  $\beta$ -glucosidases, the scores were much lower, the best ones being with human and *Ciona intestinalis* enzymes (19% identity). Sequences from the NCBI were searched to complement the set of unclassified glycoside hydrolase sequences already present in CAZy previously collected based on the bile acid-glycosidases, and integrated to create a new family, that we have designated as Glycoside Hydrolase family 116 (GH116). Once the conserved catalytic regions were aligned, a distances tree was obtained (Figure 3.5).



1.0



**Figure 3.5:** Phylogenetic tree of family GH116 using Ward hierarchical clustering distances. The leaves of the tree indicate the organism genus and species information (e.g. *Hom\_sapie* corresponds to *Homo sapiens*), the gene or Locus name, the EC activities if characterized experimentally, and a database accession number. Tree branches were colored according to identified significant subgroups using Dendroscope 2.3

## Identification of the catalytic residues of SSO1353

The identification of key active-site residues in a glycoside hydrolase is crucial to determine the catalytic machinery allowing a more precise classification of this class of enzymes (Cantarel et al. 2009). These residues can be identified by using several different techniques, site-directed mutagenesis followed by kinetic analysis of the mutants

To identify highly conserved aspartic and glutamic acid residues, we aligned the amino acid sequence of SSO1353 to eight other hypothetical proteins identified by BLAST analysis with an identity  $\geq 22\%$  in particular: the ORFs SSO2674, SSO1948 and SSO3039 from *Sulfolobus solfataricus* P2, ST2526, ST2609 and ST0868 from *Sulfolobus tokodaii*, DUF680 and DUF608 from *Caldivirga maquilingensis*. The multi-alignment led to the identification of 15 Asp/Glu residues highly conserved. Among these residues, Glu335, Asp406, and Asp462 were invariant while Asp458 is highly conserved showing a lysine in SSO3039 sequence. (Figure 4.5).

	*	#	
SSO2674	LTkdgrFAVY <b>ED</b> PFVTKLMNTIGSMTFDGL-----GFT-LLELYRDLVISA	345	
ST2526	FTkdgrFAIY <b>ED</b> PEISLLMNTIGAMTWDSA-----SFP-LLELYPDLVKKM	343	
duf680	LTkdgrLAVY <b>ES</b> LSIAPLMSTIGSMTWDGL-----SFA-LLDLFPDLTVKM	353	
<b>SSO1353</b>	LDEKGRFAIY <b>EAP</b> QNCPLYGTIG-ACYEFG-----SLP-VILMFPELEKSF	368	
SSO1948	LDEKGRFAIY <b>EAP</b> QNCPLYGTIG-TCYEFG-----SLP-VILMFPELEKLF	366	
ST2609	LDEKGRFSIF <b>EAP</b> TNFPYQGTIG-TCYEFG-----SLP-ILSFFPELDKSF	352	
DUF608	LTRDGRFSIL <b>EG</b> VEVCPCHGTLAGACYETG-----SLP-VVLMFPELEKSL	376	
ST0868	LDEKGRFGIM <b>EG</b> TQVGTMLSTIGVCYETG-----SLP-VVLMFPMLEKST	374	
SSO3039	LTkdGFFGIW <b>EG</b> YFDTSDYRKVGKYPYTGGPENTALNTIDVLLYALPGVMLLFPPELAKNI	370	
	<b>E335</b>		
	*	#	
SSO2674	PYW-----W <b>TD</b> LGPTLVLMLYRDYVFTSNREILEKNYNKIKEIIDWLI	419	
ST2526	GYP-----W <b>TD</b> LGSTWVLMYRDYKFTNDLAFLKRNRYKMKKEVIDWLI	417	
duf680	LYP-----W <b>ND</b> LGSTWILMIYRDYLLTGNVEVLRNIDKMREVIDWLI	427	
<b>SSO1353</b>	PPR-----W <b>KD</b> MNPSLILLVYRYFKFTNDIEFLKEVYPILVKVMDWEL	443	
SSO1948	PPK-----W <b>KD</b> MNPSLILLVYRYFKFTNDIDFLKEVYPTIVKVMDWEL	441	
ST2609	PPK-----W <b>KD</b> LNPTYILLIYRYKLTGDI EFLKSVYDKVKKAFEWEL	427	
DUF608	PPR-----W <b>KD</b> LNSTYILLVHRYFKRSNDVEFKEIYPKLIKAFEWVL	451	
ST0868	PPK-----W <b>KD</b> TNTTFVLMVYRYLRTKDKEFLKSVYPYVKKAMSWII	449	
SSO3039	DPKGRMPHYIRHSLTVDTYER <b>VD</b> INPEFVLLYYLIAKYTGDRFLT SVYEVARNAIESIM	490	
	<b>D406</b>		
	#	*	
SSO2674	RKdMDN-----DCIPDSKGGY <b>DN</b> -----SY <b>DG</b> THMYGASSYI	451	
ST2526	SLdKDK-----DCIPDSKGGF <b>DN</b> -----SY <b>DG</b> TYMYGASSYV	449	
duf680	SRdYDG-----DCIPDSRGGF <b>DN</b> -----SY <b>DG</b> TNMYGASSYI	459	
<b>SSO1353</b>	RQCK-----GNLPFMEGEM <b>DN</b> -----AF <b>DAT</b> I IKGHDSYT	473	
SSO1948	RQCR-----DGLPFMEGEM <b>DN</b> -----AF <b>DAT</b> I IKGHDSYT	471	
ST2609	KFSR-----YGL---EGKM <b>DS</b> -----AF <b>DV</b> TPIKGINSYT	454	
DUF608	VQDKDG-----DGVPELSGD <b>DT</b> -----GF <b>DAM</b> SVKGFDSYT	483	
ST0868	SKDKDG-----DGLPEVDGS <b>DQ</b> -----GF <b>DC</b> VP IEGVCSYI	481	
SSO3039	RTQTLDGLFYLTLP SGIEWMRHVNML <b>KV</b> SDAHKILGYHTLALSMQTL <b>DD</b> WSWLGFSPPYV	550	
	<b>D458</b>	<b>D462</b>	

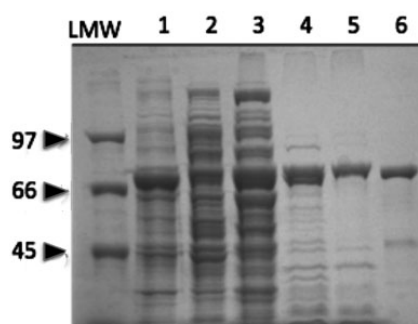
**Figure 4.5:** Multi-alignment of SSO1353 with other homologous hypothetical proteins. Highly conserved Asp/Glu residues are indicated with “#”, invariant residues are indicated with “\*” both in bold red. The codes indicate the ORF numbers: SSO2674, SSO1948, SSO3039 (*S. solfataricus* P2); ST2526, ST2609, ST0868, (*S. tokodaii*); DUF680, DUF608 (*Caldivirga maquilingensis*).

To understand the catalytic role of these conserved residues, the mutants Glu335Gly, Asp406Gly, Asp458Gly, and Asp462Gly were prepared by site-directed mutagenesis (Cobucci-Ponzano et al. 2010).

#### Expression and purification of SSO1353 wild type and mutants

SSO1353 wild type, previously cloned (for the cloning strategy see Cobucci-Ponzano et al. 2010), and its mutants were expressed and purified (Figure 5.5) as described in *Experimental Procedures*.

During this procedure all the proteins showed identical behaviour in expression level, suggesting that the mutations did not affected the stability of the enzymes and the purification steps yielded proteins with similar concentrations and purification degrees (not shown).



**Figure 5.5:** SDS-PAGE analysis of SSO1353. LMW, molecular weight markers (kDa); lane 1, *E. coli* BL21(DE3)Ril/pET1353 soluble protein extract (45 µg); lanes 2–4, protein extract after heat treatment at 55, 75, and 85°C, respectively (70, 120 and 84 µg, respectively); lane 5, typical sample after hydrophobic chromatography (21 µg); lane 6, sample after gel filtration (3 µg).

SSO1353 is a monomer of about 76 kDa in native conditions and the purified enzyme is optimally active at pH 5.5 at 65 °C, active on  $\beta$ -glucosides and  $\beta$ -xylosides (Table 1.5) with the highest specificity for MU-Glycosides. The enzymatic characterization showed that SSO1353 is a *retaining*  $\beta$ -glycosidase specific for gluco- and xylosides (EC 3.2.1.21/37) showing increased affinity for substrates having hydrophobic leaving groups (Cobucci-Ponzano et al. 2010).

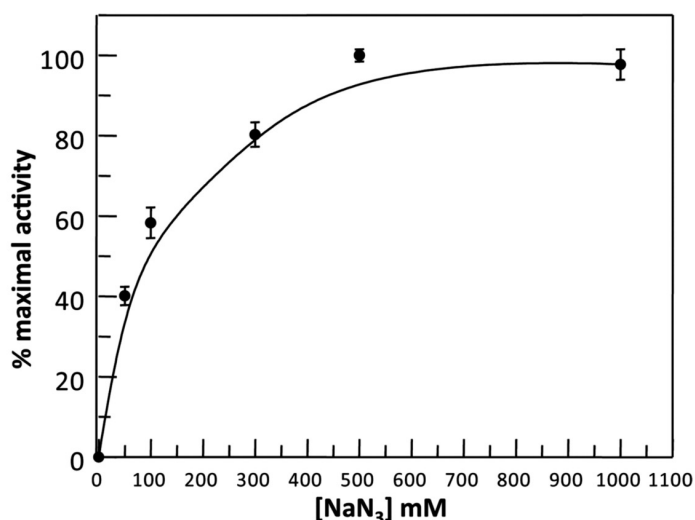
Substrate	$k_{cat}$ ( $s^{-1}$ )	$K_M$ (mM)	$k_{cat}/K_M$ ( $s^{-1} mM^{-1}$ )
2Np- $\beta$ -D-Glc	4.7 $\pm$ 0.3	12.8 $\pm$ 2.5	0.37
4Np- $\beta$ -D-Glc	4.9 $\pm$ 0.5	53.7 $\pm$ 11.8	0.09
4Np- $\beta$ -D-Xyl	4.3 $\pm$ 0.4	24.6 $\pm$ 7.8	0.17
MU- $\beta$ -D-Glc	1.2 $\pm$ 0.1	2.6 $\pm$ 0.5	0.47
MU- $\beta$ -D-Xyl	0.8 $\pm$ 0.1	1.2 $\pm$ 0.3	0.71

**Table 1.5:** Steady state kinetic constants of the recombinant SSO1353

### Reactivation of SSO1353 mutants

The mutants assayed at 65°C on 2NP-Glc 40 mM in 50 mM sodium citrate buffer pH 5.5 were completely inactive indicating that the mutations affected the catalytic machinery of SSO1353.

When 1 M sodium azide was included in the assay on 40 mM 2NP-Glc we observed the reactivation of the Asp462Gly mutant, which showed a specific activity of 0.5 U mg<sup>-1</sup>, which is about 7-fold lower than that of the wild type (3.6 U mg<sup>-1</sup>) assayed in the same conditions. Instead, the external ion did not modify the specific activity of the wild type (data not shown) and did not reactivate the Glu335Gly, Asp406Gly, and Asp458Gly mutants. The mutant Asp462Gly was assayed at standard conditions on 40 mM 2NP-Glc in the presence of increasing concentrations of sodium azide: the maximal activity was observed at 0.5 M sodium azide (Figure 6.5).



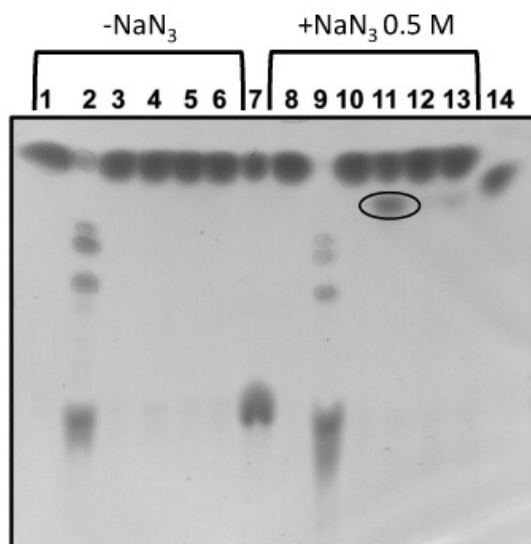
**Figure 6.5:** Dependence of activity of D462G mutant on different concentrations of sodium azide.

At these conditions the kinetic constants were  $k_{\text{cat}}$  of  $0.64 \pm 0.1 \text{ s}^{-1}$ ,  $K_{\text{M}}$  of  $16.2 \pm 6 \text{ mM}$ , and  $k_{\text{cat}}/K_{\text{M}}$  of  $0.04 \text{ s}^{-1} \text{ mM}^{-1}$ , showing that Asp462Gly maintained a similar affinity for the substrate, but a specificity constant 10-fold lower than the wild type in standard conditions (Table 2.5).

	Wild type	D462G + 0.5 M NaN <sub>3</sub>
$k_{\text{cat}} (\text{s}^{-1})$	$4.7 \pm 0.3$	$0.64 \pm 0.1$
$K_{\text{M}} (\text{mM})$	$13 \pm 2$	$16.2 \pm 6$
$k_{\text{cat}}/K_{\text{M}} (\text{s}^{-1} \text{ mM}^{-1})$	0.37	0.04

**Table 2.5:** Comparison between kinetic constants of SSO1353 wild type and D462G mutant with 0.5 M sodium azide.

Reaction mixtures were analyzed by TLC: the Asp462Gly produced a novel compound (highlighted in Figure 7.5), which was observed only in trace amounts with the other mutants. Instead, the wild type completely converted the substrate producing transglycosylation products (Figure 7.5).



**Figure 7.5:** TLC analysis of the reaction mixtures of SSO1353 wild type and mutants. Standard assays were performed overnight at 65°C on 40 mm 2Np-Glc by using 11 µg of enzyme. Lane 1, blank with no enzyme; lane 2, wild type; lane 3, E335G mutant; lane 4, D462G; lane 5, D458G; lane 6, D406G; lane 7, standards (2Np-Glc and Glc); lane 8, blank with no enzyme; lane 9, wild type; lane 10, E335G; lane 11, D462G; lane 12, D458G; lane 13, D406G; lane 14, β-glucosyl-azide standard.

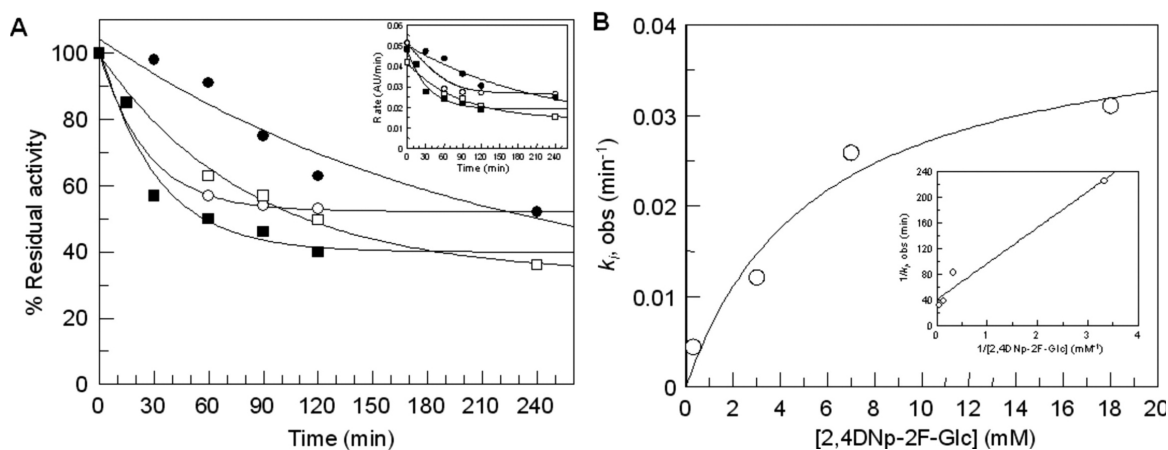
Asp462Gly reaction mixtures in preparative scale allowed the isolation and structural characterization of this product (performed by Prof. M.M. Corsaro of University of Naples “Federico II”) that was unequivocally identified as β-glucosyl azide (β-GlcN<sub>3</sub>). The reactivation in the presence of the external ion and the anomeric configuration of this product strongly indicate that Asp462 is the acid/base of the reaction (for details on the *chemical rescue* see the General Introduction).

#### Identification of the catalytic nucleophile of SSO1353

To identify the nucleophile of the reaction we used an approach combining a mechanism-based inhibitor approach and nano-electrospray ionization tandem mass spectrometry (nano-ESI-MS/MS).

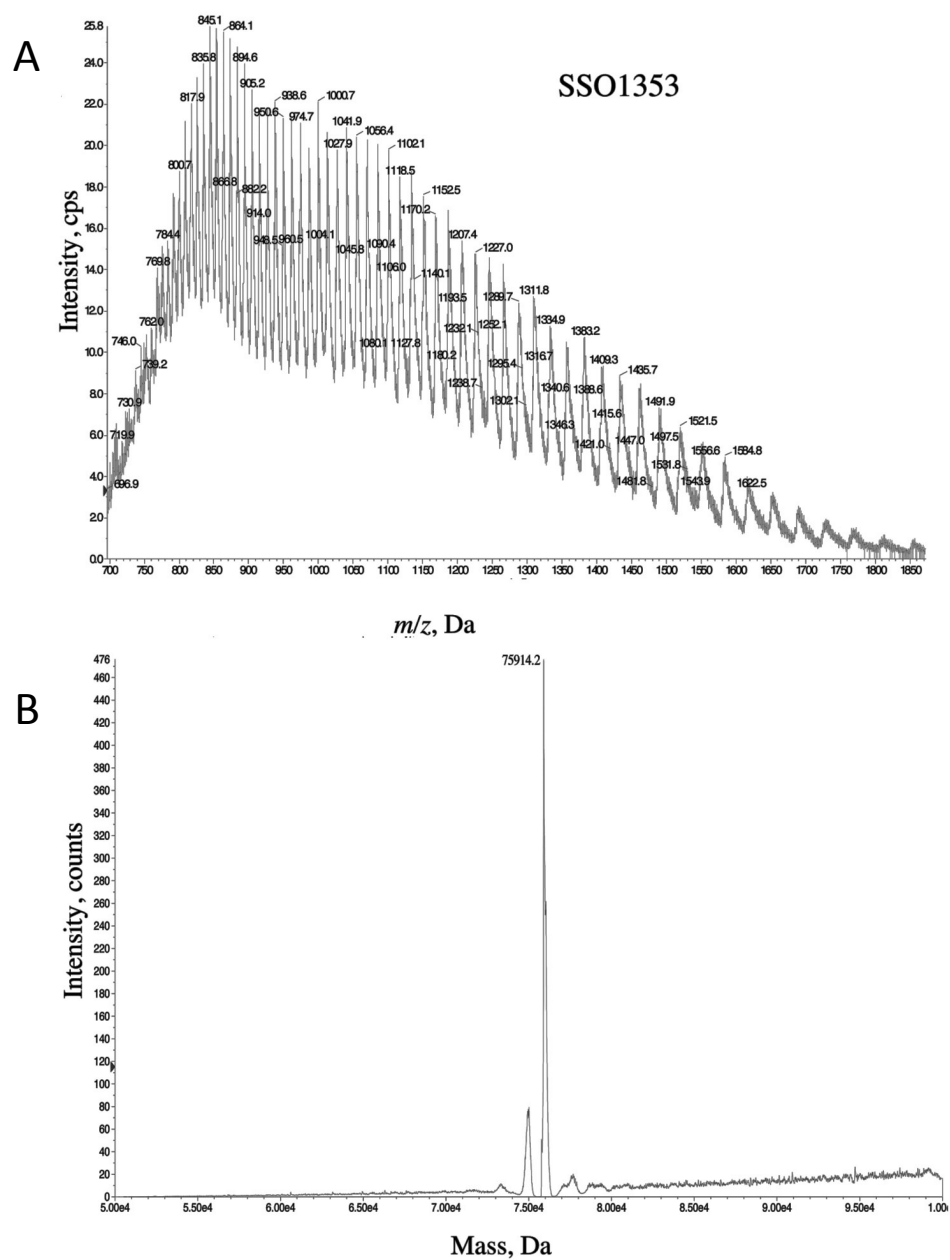
Time dependent inactivation of SSO1353 was observed upon incubation of the enzyme with 2,4DNP-2F-Glc (Figure 8.5A,B). Inhibition was incomplete after 4 h of incubation (about 40%) even at the highest concentration of inhibitor used (18 mM).

This is not surprising because GHs inactivated by mechanism-based inhibitors are catalytic competent and the release of the inhibitor, occurring via turnover of the intermediate via hydrolysis or transglycosylation, has been well documented (Williams & Withers, 2000). At these conditions we obtained the following inactivation parameters:  $k_i = (6.9 \pm 1.3) \times 10^{-4} \text{ s}^{-1}$ ;  $K_i = 5.5 \pm 2.7 \text{ mM}$ ;  $k_i/K_i = 1.2 \times 10^{-4} \text{ s}^{-1} \text{ mM}^{-1}$ .

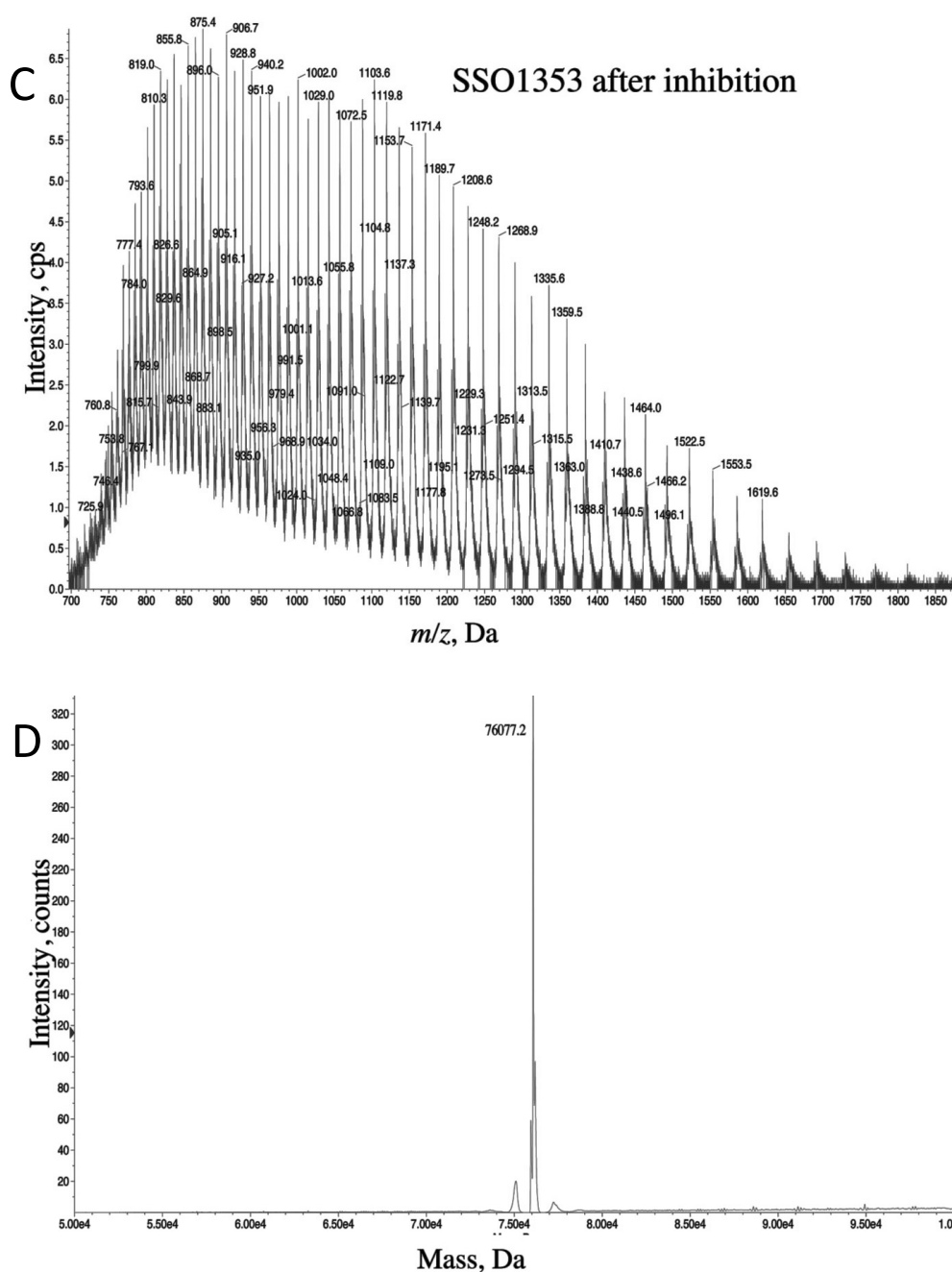


**Figure 8.5:** Time-dependent inactivation of SSO1353 using 2,4DNp-2F-Glc inhibitor. (A) plot of % residual activity versus time at four 2,4DNp-2F-Glc concentrations (● 0.3 mM, □ 3 mM, ○ 7 mM, and ■ 18 mM) with the plot of rate versus time shown as inset.  $k_i, \text{obs}$  were obtained from fitting the curves in (A) to a single exponential decay with offset because time-dependent inactivation did not decay to zero, and plotted versus inhibitor concentration (B), to determine  $K_i$  and  $k_i$ . The reciprocal plot is shown as inset in (B).

SSO1353 samples incubated in the absence and the presence of 2.9 mM 2,4DNp-2F-Glc for 2 h were analyzed by single-stage nano-ESI-MS in order to monitor alteration in the molecular mass of the protein. Nano-ESI mass spectra of intact proteins yield series of multiply charged molecular ion peaks with 40-100 positive charges under the experimental condition applied. Molecular mass of SSO1353 in the absence of inhibitor was measured to be  $75914 \pm 6 \text{ Da}$ , which is comparable to the theoretical average molecular mass ( $75907.7 \text{ Da}$ ) within experimental error (0.009%) (Figure 9.5).



**Figure 9.5:** Single-stage nano-ESI (A) and corresponding deconvoluted (B) mass spectra acquired on SSO1353 samples incubated in absence of 2,4DNp-2F-Glc. Multiply charged molecular ion peaks with 40-100 positive charges were observed under the experimental condition applied.



**Figure 10.5:** Single-stage nano-ESI (C) and corresponding deconvoluted (D) mass spectra acquired on SSO1353 samples incubated in presence of 2,4DNp-2F-Glc. Multiply charged molecular ion peaks with 40-100 positive charges were observed under the experimental condition applied. After inhibition, molecular ion peaks shift towards higher  $m/z$  values leading  $163 \pm 5.5$  Da shift in molecular mass (D) respecting to that observed in the absence of inhibitor (Figure 9.5B)

After inhibition, molecular ion peaks shift towards higher  $m/z$  values leading a molecular mass of  $76077 \pm 5$  Da and accounting for  $163 \pm 5.5$  Da difference between the two species (Figure 10.5). To gain further evidence and a more detailed structural insight into the site-directed inhibition, samples were proteolytically digested by pepsin and the resulting peptide mixtures were analyzed by nano-HPLC-ESI-MS/MS in IDA mode. Based on MS/MS sequence data, 91% and 87% protein sequence



coverage were respectively obtained in the absence and in the presence of inhibitor (data not shown).

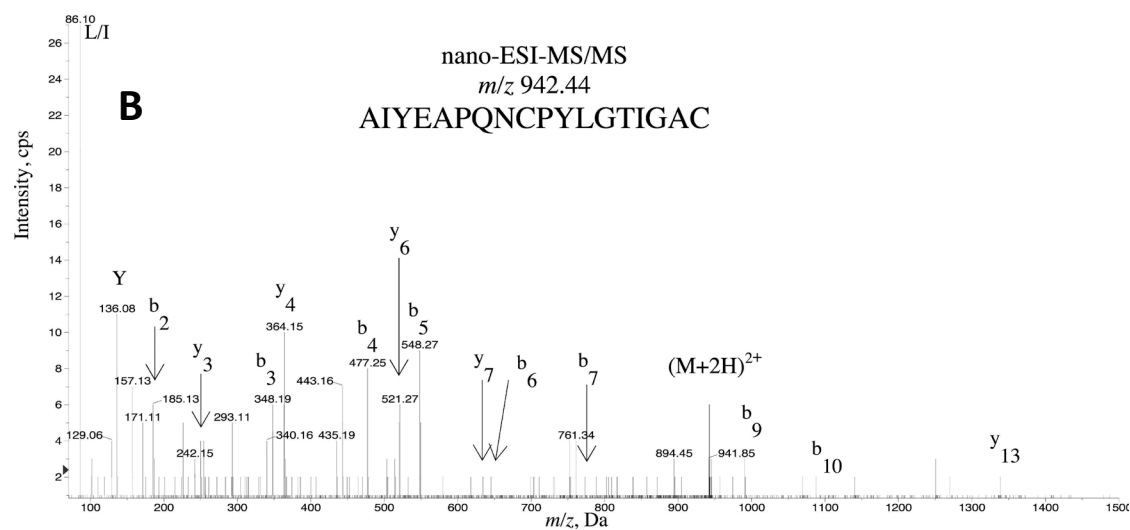
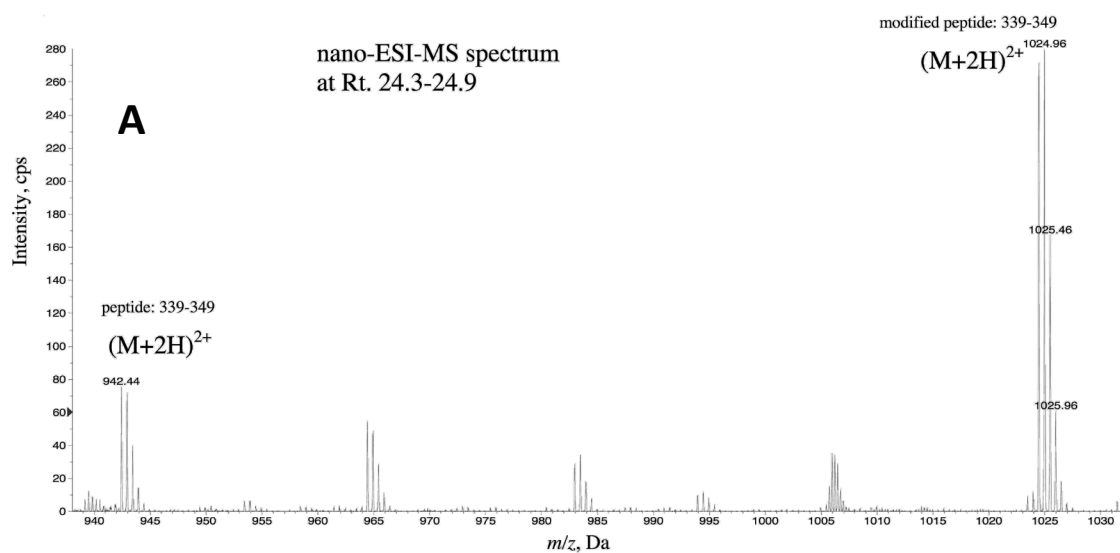
Interestingly, in the inhibited sample five peptides comprising residues 332-345, 332-343, 332-347, 332-348 and 332-349 showed considerable decrease in intensity, and in the same time, six new peptide molecular ions appeared in the corresponding scans (Table 3.5).

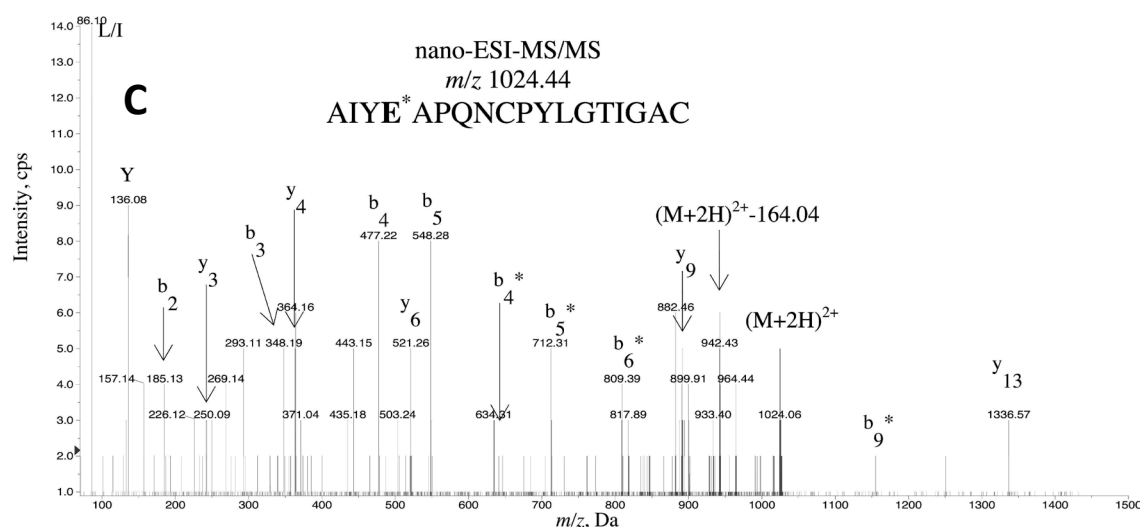
From-To	Unmodified and modified peptide sequences	Mw (calc.)	SSO1353			SSO1353 inhibited		
			Rt (min)	m/z	Intensity (cps)	Rt (min)	m/z	Intensity (cps)
332-345	AIYEAPQNCPYLGT	1538.708	20.7	770.36	26	21.2	770.36	10
	AIYE*APQNCPYLGT	1703.188	-	n.d.	-		852.39	34
332-343	AIYEAPQNCPYL	1380.638	21.8	691.33	30	22.0	691.33	17
	AIYE*APQNCPYL	1544.689	-	n.d.	-		773.35	44
332-347	AIYEAPQNCPYLGITIG	1708.813	23.2	855.42	42	23.2	855.42	16
	AIYE*APQNCPYLGITIG	1873.293	-	n.d.	-		937.44	37
332-348	AIYEAPQNCPYLGITIGA	1779.85	23.5	890.94	280	23.7	890.93	63
	AIYE*APQNCPYLGITIGA	1943.898	-	n.d.	-		972.96	240
332-349	AIYEAPQNCPYLGITIGAC	1882.859	24.8	942.46	320	24.6	942.44	106
	AIYE*APQNCPYLGITIGAC	2046.07	-	n.d.	-		1024.46	384
332-340	AIYEAPQNC	1007.447	-	n.d.	-	25.2	n.d.	-
	AIYE*APQNC	1171.495	-	n.d.	-		586.76	94

**Table 3.5:** Characteristic peptide molecular ions containing residue E at position 335 observed during nano-HPLC-ESI-MS/MS IDA analyses of SSO1353 incubated in the absence and in the presence of 2,4-dinitro-phenyl-2-deoxy-2-fluoro-glucopyranoside (2,4DNP-2F-Glc) inhibitor and digested by pepsin. Peptide sequences (both unmodified and modified) were elucidated by the interpretation of nano-ESI-MS/MS spectra acquired on the doubly charged ( $z=2$ ) precursor ions (Figure 11.5A,B,C). Modification corresponds to the covalent attachment of 2,4DNP-2F-Glc ligand at E335 (indicted in the sequence as E\*).

The peptide molecular ion pairs corresponding to the normal and the modified sequences eluted at the same retention time and thus they were detected in the same survey scan in the sample containing 2,4DNP-2F-Glc. Therefore, they are likely due to in-source ion fragmentation process indicating a relatively labile bond between amino acid and inhibitor. These peptides showed an increase of 164.05 Da in molecular mass which corresponds well to the difference between unmodified and 2F-Glc modified peptides, and indicated that the ligand was likely bound to one of the amino acids present in peptide 332-349.

To confirm the site of modification, nano-ESI-MS/MS spectra of the unmodified/modified peptide pairs were analyzed (Table 3.5, Figure 11.5A,B,C).





**Figure 11.5:** Identification of the catalytic nucleophile site of SSO1353 by nano-HPLC-ESI mass spectrometry. (A) Nano-ESI mass spectrum of peptides eluted at 24.3–24.9 min of SSO1353 incubated with 2,4DNp-2F-Glc for 2 h. Doubly charged peptide ions at  $m/z$  942.44 and 1024.46 correspond to the unmodified and the modified peptide 339–349, respectively. Tandem mass spectra on the unmodified (B) and modified peptides 339–349 (C) reveal modification on E4 residue.

MS/MS spectra show a very similar fragmentation pattern yielding characteristic b type N-terminal fragment ions at the low  $m/z$  range. Based on these ions, and in particular, on the appearance of  $b_n^*$  ( $n \geq 3$ ) modified fragment ions at  $m/z$  641.28 ( $b_3^*$ ), 712.32 ( $b_4^*$ ) and 809.37 ( $b_5^*$ ) in the inhibited sample, modification was unequivocally localized on amino acid Glu335. Therefore, it was concluded that Glu335 is the nucleophile of the reaction of SSO1353.

## **Chapter III**

### **Discussion**

In the laboratory in which I have performed my thesis, we cloned, expressed and characterized the product of the gene *sso1353* from the hyperthermophilic archaeon *S. solfataricus*. The molecular characterization revealed the specificity of the enzyme for gluco- and -xylosides  $\beta$ -bound to hydrophobic groups that are hydrolyzed by following a *retaining* reaction mechanism (Cobucci et al. 2010). In addition, site-directed mutagenesis of conserved glutamic/aspartic amino acids and the chemical rescue of the  $\beta$ -glycosidase activity of the mutants, combined with the use of mechanism based inhibitors and mass spectrometric analysis, allowed us to identify Asp462 and Glu335 as the acid/base and the nucleophile of the reaction, respectively. Amino acid sequence analysis showed that SSO1353 shared identity with other hypothetical proteins and, remarkably, with eukaryotic non-lysosomal bile acid  $\beta$ -glucosidases.

So far, SSO1353 was not assigned to a defined glycoside hydrolase family in the carbohydrate active enzyme database. On the basis of our findings we propose that SSO1353 and its homologs define a new sequence-based family, namely GH116, which presently includes enzymes with  $\beta$ -glucosidases (EC 3.2.1.21),  $\beta$ -xylosidases (EC 3.2.1.37), or glucocerebrosidases (EC 3.2.1.45) activity. As for the other GH families, the retaining reaction mechanism and the catalytic role for the acid/base and the nucleophile, experimentally determined here, can be easily extended to all the enzymes belonging to this new family.

Interestingly, all the archaeal putative enzymes belonging to this new family are from Crenarchaea, and the vast majority originates from the genus *Sulfolobus*. A PSI-BLAST search conducted using SSO1353 as the query sequence retrieved (with low scores) uncharacterized bacterial glycosidases belonging to families GH15, GH63, and GH78. The latter families include mainly glucoamylases,  $\alpha$ -glucosidases, and  $\alpha$ -L-rhamnosidases, respectively, and are characterized by an  $(\alpha/\alpha)_6$  fold. Although SSO1353 is inactive on  $\alpha$ -glycosides, this perhaps hints at structural similarities with enzymes from family GH116. Similar structural similarity between  $(\alpha/\alpha)_6$  fold glycoside hydrolase families degrading both  $\alpha$  and  $\beta$  glycosidic bonds have already been described (Stam et al. 2005).

The phylogenetic analysis (Figure 3.5) shows that sequences from the new family GH116 can be subdivided into two major groups, one containing sequences from Archaea and another one composed mostly of sequences from Cyanobacteria and Eukaryotes. The archaeal subgroup can be further subdivided into at least two subgroups, in which, interestingly, all the archaeal homologs of SSO1353 are present as multiple copies in the genomes of *Caldivirga maquilingensis*, *S. tokodaii*, *S. solfataricus*, and in the six strains of *S. islandicus*. The *sso1353* homologs with identity >80%, lie downstream of genes encoding endoglucanases (Maurelli et al. 2008), and, interestingly, in *S. solfataricus*, this gene arrangement occurs twice. Presumably, the  $\beta$ -glycosidase activity of SSO1353 is involved, in combination with the secreted endoglucanase, in the degradation of exogenous glucans used as carbon energy source or, possibly, of the exo-polysaccharides (EPS) that are produced by *S. solfataricus* itself (Zolghadr et al. 2010; Elferink et al. 2001). These analysis indicate SSO1353 as an interesting candidate to be utilized in the biofuel industry, in combination with endoglucanase, to hydrolyze vegetable raw materials to obtain simple fermentable sugars.

Other sso1353 homologs, with identities in the range 21–33% exemplified by sso2674 and sso3039 in *S. solfataricus*, flank a putative peptidase or a putative gluconolactonase, respectively. These two other subgroups of enzymes similar to SSO1353 are present also in *C. maquilingensis*, *S. tokodaii*, and *S. islandicus* showing a remarkable identity (>80%) within each subgroup. The observation that the archaeal  $\beta$ -glycosidases from this novel GH family can be subgrouped according to their identity suggests they are present in multiple copies for functional purposes, possibly, for the degradation/modification of different substrates. A more detailed characterization of these enzymes is needed to understand their function *in vivo*.

The other major subdivision of the family is prone to be subdivided into several subgroups, one containing sequences from Cyanobacteria, the other having plant, animal, and mixed bacterial subdivisions. One of the members of the animal subgroup in this newly proposed family is human non-lysosomal glucosylceramidase or  $\beta$ -glucosidase 2 (GBA2). This enzyme, previously described as bile acid  $\beta$ -glucosidase (H Matern et al. 2001), is involved in the catabolism of glucosylceramide, which is then converted to sphingomyelin (Boot, et al 2007). Glucocerebrosidases are important enzymes involved in the metabolism of gangliosides and globosides. Deficiency of this enzymatic activity is the cause of the most common lysosomal storage disorder named Gaucher disease (Grabowski 2008) resulting from a defect in the lysosomal acid  $\beta$ -glucosidase (GBA1) belonging to GH30. This deficiency leads to the accumulation of glycosylceramides in certain organs, typically spleen, kidney, lungs, brain, and bone marrow (Grabowski et al. 1990). The finding that other cell types of Gaucher patients did not show accumulation of glycosylceramides suggested the existence of an alternative catabolic pathway that later was demonstrated to be catalyzed by GBA2 (Boot, Verhoek, Donker-Koopman, Anneke Strijland, van Marle, Overkleeft, Wennekes, and J. M. F. G. Aerts 2007b). This enzyme is ubiquitously expressed and it is associated to the cell surface. GBA2 is inactive on MU-Xyl, is inhibited by hydrophobic deoxynojirimycin (DNJ), and it is relatively insensitive to CBE (Boot, Verhoek, Donker-Koopman, Anneke Strijland, van Marle, Overkleeft, Wennekes, and J. M. F. G. Aerts 2007b; H Matern et al. 2001). In humans, no known pathologies related to defects of GBA2 have been reported so far while only in certain mice strains treatments with NB-DNJ or *gba2* gene knock-outs led to impaired spermatogenesis (van der Spoel et al. 2002). However, such deleterious effects were not observed in other organisms including humans (Wilhelm Bone et al. 2007; Boot et al. 2007; Amory et al. 2007).

These studies demonstrate the importance of understanding at the molecular level the reaction mechanism and the catalytic machinery of carbohydrate active enzymes for the development of specific inhibitors for bio-medical applications. The experimental identification of the catalytic amino acids of SSO1353 reported here, allows to easily identifying the catalytic machinery of human GBA2 despite the low sequence identity (18%) between the two enzymes. In GBA2 the nucleophile and the acid/base of the reaction are Glu528 and Asp678, respectively, which, as observed in a multi-alignment of putative glucocerebrosidases from mammals, plants, and tunicates belonging to this new GH family, are located in two conserved motifs (Figure 12.5).

Pan	-----LPEELG-RNMCH-----LRPTLRDYGRFGYLEGQ	456
Homo	-----LPEELG-RNMCH-----LRPTLRDYGRFGYLEGQ	529
Ciona	-----NGKVPETLQNTTKVD-----SCDFLKEYGKFAYLEGQ	498
Arabidopsis	GLKNDIDVPHQNDTAVSVLEKMASTLEELHASTTSNSAFGTKLLEEGEENIGHFLYLEGI	535
Sulfolobus	-----EKGRFAIYEAP	337
	: *: *	
Pan	EYRMYN--TYDVHIFYASFALIMLWPKLELSLQYDMALATLREDLTRRRYLMSGVMAPVKR	514
Homo	EYRMYN--TYDVHIFYASFALIMLWPKLELSLQYDMALATLREDLTRRRYLMSGVMAPVKR	587
Ciona	EYKMYN--TYDVHIFYASIALAYLWPKLELSVQYDIATSIHLSNPQPHKYLMDGVTAPVKT	556
Arabidopsis	EYRMWN--TYDVHIFYASFALVMLFPKLELSIQRDFAAAVMLHDP TKVKT LSEGQWVQRKV	593
Sulfolobus	QNCPLYGTIGACYEFGSLPVLMPFELEKSFLKLLIRHIRED-----	379
	: : : : : : : : : : : : : : : :	
Pan	RNVIPHDIGDPDDEPWLRVNAYLIHDTADWKDLNLKFVLQVYRDYYLTGDQNFLKDMWPV	574
Homo	RNVIPHDIGDPDDEPWLRVNAYLIHDTADWKDLNLKFVLQVYRDYYLTGDQNFLKDMWPV	647
Ciona	PNVVPDVGCPEDPWLRVNTYFVHDTADWRDLNPKFVLQAYRDYYITKDIDFLKAMWPI	616
Arabidopsis	LGAVPHDLG--INDPWFEVNGYTLHNTDRWKDLNPKFVLQVYRDVVATGDKKFASAVWPS	651
Sulfolobus	-GYVPHDLGYHSLDSPIDG----TTSPPRWKDMNPSLILLVYRYFKFTNDIEFLKEVYPI	434
	. : : : : : : : : : : : : : : *	
Pan	CLAVMESEMKFDDKDHGLIENGGYADQTYDGWVTTGPSAYCGGLWLA AVAVMVQMAALCG	634
Homo	CLAVMESEMKFDDKDHGLIENGGYADQTYDGWVTTGPSAYCGGLWLA AVAVMVQMAALCG	707
Ciona	CKIVMEQSMRHDKDNDGLIENSGAADQTFDGWCVTGPSAYCGGLWLAALRCMEEAADILH	676
Arabidopsis	VYVAMAYMAQFDKDGDMIENEGFPDQTYDTWSASGVSA YCGGLWVAALQAASALARVVG	711
Sulfolobus	LVKVMDWELRQCKGNLPFMEG--EMDNAFTATIKGHDSYTSLSLFIAMREIAKLVG	492
	. * : * . : : . * : : * . * . * . * : : : *	

**Figure 12.5:** Multi-alignment of SSO1353 with glucosylceramidases. Invariant residues are indicated with “\*”; increased level of conservation is indicated with “.” and “.”. The residues corresponding to the nucleophile Glu335 and acid/base Asp462 of SSO1353 are boxed. Pan is XP\_001167952.1 from *Pan troglodytes*; Homo is NP\_065995.1 from *Homo sapiens*; Ciona is XP\_002127036.1 from *Ciona intestinalis*; Sulfolobus is SSO1353 from *S. solfataricus* P2. (Cobucci Ponzano et al. 2010)

In particular, amino acids with hydrophobic side chains are almost invariant in the position preceding the catalytic glutamic and aspartic acids in the enzymes belonging to the new family GH116 (Figure 12.5). Our findings can now allow the planning of more detailed site-directed mutagenesis studies to better understand the molecular bases of the substrate recognition of GBA2.

SSO1353 has substrate specificity and inhibitor sensitivity slightly different from those of GBA2. In fact, the archaeal enzyme can hydrolyze both aryl  $\beta$ -gluco and  $\beta$ -xylosides and it is inhibited with mM affinity by both NB-DNJ and CBE. Instead, GBA2 is inactive on MU-Xyl and it is relatively insensitive to CBE (Boot et al. 2007). These differences presumably reflect the different function of the two enzymes *in vivo*: the wider substrate specificity of the archaeal enzyme might allow to degrade a variety of substrates ensuring an efficient availability of sugars as energy source while GBA2 is involved in a well defined catabolic pathway. The purification of GBA2 is made difficult by its instability to detergents precluding its production in abundant and homogeneous form (Boot et al. 2007). Instead, robust GBA2 homologs from hyperthermophilic Archaea can be more easily expressed and purified from conventional hosts allowing more simple structural studies that might be easily extended to the human counterpart.

## **Chapter IV: A $\alpha$ -mannosidase from *Sulfolobus solfataricus***

### **Introduction**



The most prevalent post-translational modification in Eukarya is glycosylation, over two thirds of all proteins in the SWISS-PROT database contain glycosylation sites, and, in humans, 1-2% of the genes are involved in the regulation of N- and O-glycosylation. When glycosylation is inhibited, the mainly observed effect is the production of misfolded and aggregated proteins that fail to reach a functional state (Helenius 1994). The importance of the added glycans varies between proteins and depends on the physiological context. Some proteins are completely dependent on glycosylation, whereas many display no dependence at all. Some become temperature-sensitive for folding, some are glycan-dependent only in one cell type but not in another and some proteins have multiple glycosylation sites and may happen that some are more important than others for its stability implying that oligosaccharide appendices have local effects on protein folding.

Trimming by glycoside hydrolases of the N-linked glycans also plays a role in the sorting process leading to glycoprotein degradation in the ER. Proteins that fail to reach their native conformation in the ER are selectively eliminated by ER-associated degradation. This fate is shared by misfolded and mutant proteins, by orphan subunits of oligomers, and by some heterologously expressed proteins. Because misfolded side products are common even under unstressed conditions, ER-associated degradation has a central clearance function in the cell (Kiser et al. 2001). When trimming by ER-mannosidase I is prevented by inhibitors or genetic manipulation, the degradation of glycoproteins essentially stops (Elbein 1991). This mannosidase removes a single  $\alpha$ -1,2-linked mannose residue from the  $\alpha$ -1,3 branch of the core oligosaccharide, resulting in a  $\text{Glc}_{0-3}\text{Man}_8\text{GlcNAc}_2$  structure. It is apparent that the resulting  $\text{Man}_8$  structures serve as part of the signal needed for ER-associated degradation. However removal of the mannose is not sufficient because most proteins that have folded normally are mannose-trimmed before leaving the ER. How the system works is not clear but this observations suggest a kind of feedback against premature degradation to the most recently synthesized glycoproteins (Reiss et al. 1996).

#### N-glycosylation in Archaea

N-linked glycosylation is an important post-translational modification that play a key role in several cell activities as the proteins stability, compartmentalisation and degradation.

Once believed as a unique prerogative of Eukaryotes it is now known that protein glycosylation is also present in prokaryotes. In particular, this post-translational modification in Archaea is a topic little explored, but it is arousing high interest. In this domain, the glycosylation process is considered a simpler version of eukaryal N-glycosylation.

In brief, Archaea exploit dolichol-phosphate like Eukaryotes instead of the bacterial undecaprenyl-pyrophosphate lipid carrier. In addition, the single subunit oligosaccharide transferase in Archaea is the simplified version of the homologous multimeric eukaryal counterpart. Nevertheless, many aspects of protein glycosylation in Archaea are obscure: the glycosylation mechanism, the complete set of genes involved and their function, and the structure of the glycosidic component are mostly unknown. However,  $\alpha$ -mannose has been frequently identified in glycosylated proteins and in exopolysaccharides in several Archaea including *S. solfataricus*,

*Methanosarcina acetivorans* and *M. maizei* (Abu-Qarn et al. 2008). Therefore,  $\alpha$ -mannosidases from this source could be involved not only in the catabolism of exogenous glycans, but also in the maturation of nascent glycoproteins.

For this reason, recently, we have characterized a new  $\alpha$ -mannosidase (Ss $\alpha$ -man) from the hyperthermophilic Archaeon *S. solfataricus* (Cobucci Ponzano et al. 2010).

### Classification of $\alpha$ -mannosidases

The  $\alpha$ -mannosidases are classified in two different classes, named class I and class II.

The class I enzymes are grouped in the GH47 family and they have a  $(\alpha/\beta)_7$  fold, are typically inhibited by 1-deoxymannojirimycin, perform hydrolysis by *inverting* the configuration of the anomeric carbon in the product if compared to the substrate, and are specific for the  $\alpha$ -(1,2) bonds in Man<sub>9</sub>GlcNAc<sub>2</sub> (EC 3.2.1.113). This family includes only bacterial and eukaryotic enzymes and are involved in early secretory pathway in ER and Golgi, where they perform N-glycans maturation and quality control by producing high-mannose Man<sub>5</sub>GlcNAc<sub>2</sub> (Mast et al. 2006).

Class II  $\alpha$ -mannosidases, belonging to GH38, follow a *retaining* reaction mechanism and are typically inhibited by swainsonine (Dorling et al. 1980). The eukaryotic enzymes from this family are from Golgi, cytosol, and lysosome.

Golgi  $\alpha$ -mannosidases are of biomedical interest as potential anti-cancer targets (Granovsky et al. 2000). Indeed, in breast, colon and skin cancers, the unusual quantitative distributions of complex carbohydrate structures, on the cell surface, are associated with disease progression and metastasis. This altered distribution is associated with abnormalities in the N-glycosylation while inhibition of key enzymes in this pathway has shown clinical potential in cancer treatment. The most extensively studied Golgi  $\alpha$ -D-mannosidase II is the enzyme from *Drosophila melanogaster* (dGMII).

dGMII acts on  $\alpha$ -(1,3), and  $\alpha$ -(1,6) mannosidic bonds, which are cleaved sequentially in the same catalytic site. This enzyme requires the presence of the terminal  $\beta$ -(1,2)-GlcNAc and is involved in the maturation/diversification of hybrid N-glycans, converting GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> into GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub>. Interestingly, dGMII shows three sugar binding sites named *catalytic*, *holding*, and *anchor* sites, occupied by Man<sub>5</sub> ( $\alpha$ -(1,6)-Man), Man<sub>4</sub> ( $\alpha$ -(1,3)-Man), and GlcNAc<sub>3</sub> ( $\beta$ -(1,2)-GlcNAc), respectively, with an essential zinc atom involved in both substrate binding and catalysis (Shah et al. 2008).

GH38 includes other Golgi special  $\alpha$ -D-mannosidases III, which are also specific for  $\alpha$ -(1,3), and  $\alpha$ -(1,6) mannosidic bonds, but are inactive on GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>. They convert the high-mannose N-glycan Man<sub>5</sub>GlcNAc<sub>2</sub> into Man<sub>3</sub>GlcNAc<sub>2</sub> in an alternate route for the production of hybrid GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub> (Kawar et al. 2001).

Lysosomal GH38  $\alpha$ -D-mannosidases, optimally active at acidic pH, hydrolyse all  $\alpha$ -mannosidic linkages on mannose glycans originating from glycoprotein catabolism to yield Man<sub>1</sub>GlcNAc<sub>(1-2)</sub> (Lal 1996). Deficiency in lysosomal  $\alpha$ -D-mannosidases causes  $\alpha$ -mannosidosis, a severe genetic disease producing progressive mental retardation in approximately 1 of 500,000 live births (Malm et al. 2008). The crystallographic

study of bovine lysosomal  $\alpha$ -mannosidase, structurally similar to dGMII, identified the molecular basis of some mutations causing mannosidosis and an interesting mechanism of activation at low pH (Heikinheimo et al. 2003).

So far, archaeal  $\alpha$ -mannosidases have been identified only in family GH38 and the only enzyme characterized of this family is the  $\alpha$ -mannosidase of the extreme acidophilic Euryarchaeon *Picrophilus torridus* (ManA). The enzyme might be involved in the utilization of exogenous glycans and in the turnover of its own glycoconjugates (Angelov et al. 2006). Moreover there have been no reports on the purification of native  $\alpha$ -D-mannosidases from Archaea.

#### The Ss $\alpha$ -man from *S. solfataricus*

Recently we have identified the gene encoding a new *retaining*  $\alpha$ -D-mannosidase (Ss $\alpha$ -man), belonging to GH38, from the Crenarchaeon *S. solfataricus*, cloned the gene and characterized both the recombinant (rSs $\alpha$ -man) and the native enzyme (nSs $\alpha$ -man).

The Ss $\alpha$ -man, encoded by the ORF Sso3006, is a trimer of 363 kDa whose activity, dependent on a single Zn<sup>2+</sup> ion per subunit, is inhibited by swainsonine, a potent inhibitor of Golgi alpha-mannosidase II.

Ss $\alpha$ -man is active in a rather wide range of pHs with maximal activity between pH 5.0 e 6.5. In addition, as expected for an enzyme from an extreme thermophilic microorganism, nSs $\alpha$ -man was optimally active at temperatures increasing 85°C. Instead, the thermal stability, though higher than that of mesophilic enzymes, was lower than that of other glycoside hydrolases from the same source showing an half-life of 1.5 min at 80 °C (Cobucci-Ponzano et al. 2010).

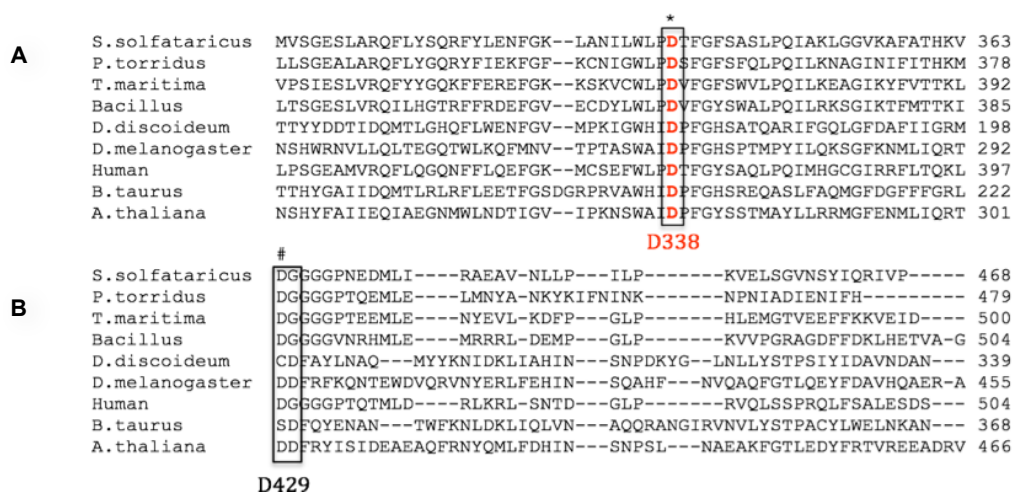
## **Chapter IV**

### **Results**

## Characterization of Ss $\alpha$ -man

### Identification of the catalytic residues of Ss $\alpha$ -man

The nucleophile residue of GH38  $\alpha$ -mannosidases has been experimentally identified in the  $\alpha$ -mannosidases from jack bean, bovine kidney (Numao et al. 2000), and fruit fly (Numao et al. 2003). The multiple amino acid sequence alignment of GH38  $\alpha$ -mannosidases allowed us to predict that in Ss $\alpha$ -man Asp338 and Asp429 are the nucleophile and the acid/base of the reaction, respectively (Figure 1.6). The acid/base Asp341 in *D. melanogaster* was identified experimentally by site-directed mutagenesis; however, the assignment of this residue in GH38 enzymes is uncertain as it falls in a region with low sequence identity (Figure 1.6). Instead the Ss $\alpha$ -man Asp338 is invariant.



**Figure 1.6:** Multiple sequence alignment of GH38  $\alpha$ -mannosidases – The regions corresponding to the catalytic nucleophile (bold and red) and the acid/base of the reaction (Ss $\alpha$ -man numbering), shown in (A) and (B), respectively, were extracted from a sequence alignment performed with the program T-Coffee (Notredame et al. 2000) by using the default settings. The catalytic residues are boxed, the symbols “\*” and “#” indicate identical and semi-conserved residues, respectively. In (B) the acid/base residue is poorly conserved in GH38, and the box encompasses two residues to include the Asp341 in *D. melanogaster* (Numao et al. 2003).

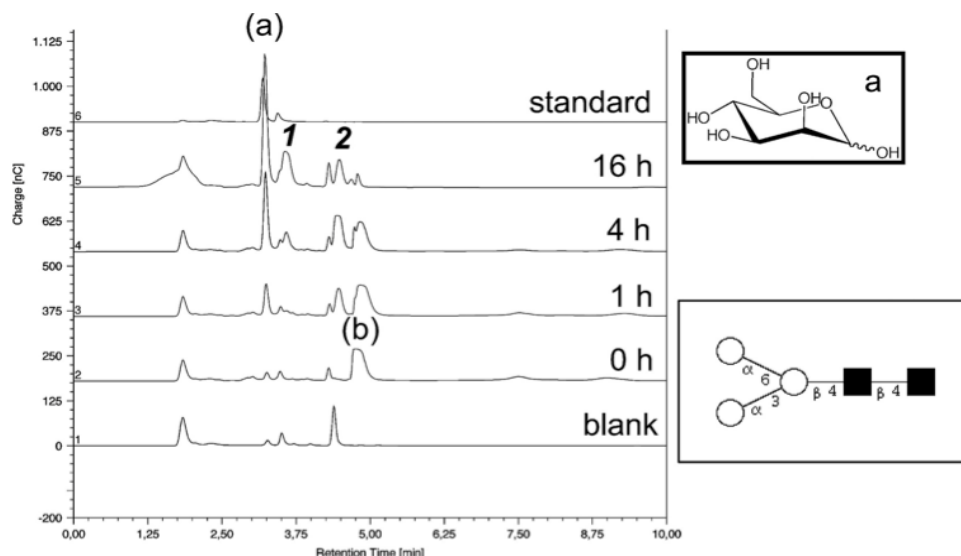
To proof the function of Asp338, the mutant Ss $\alpha$ -man Asp338Gly, previously prepared, was expressed and purified from *E. coli*. The mutant showed no activity as expected for an enzyme in which an essential catalytic residue was deleted. In addition, when assayed at standard conditions and in the presence of 1 M sodium formate as external nucleophile, the activity of the enzyme was partially rescued ( $3.5 \times 10^{-4} \pm 0.9 \times 10^{-4}$  U/mg vs.  $2.5 \pm 0.7$  U/mg for the wild type enzyme (Cobucci-Ponzano et al. 2010).

Unfortunately, no trans-mannosylation product could be observed by thin layer chromatography analysis of the reaction mixtures, precluding the possibility of determine the structure of the products; however, the multi-alignment, the inactivation by mutation, and the chemical rescue strongly indicate that Asp338 is the nucleophile of the reaction of Ss $\alpha$ -man.

### Activity of Ss $\alpha$ -man on high-mannose N-linked oligosaccharides

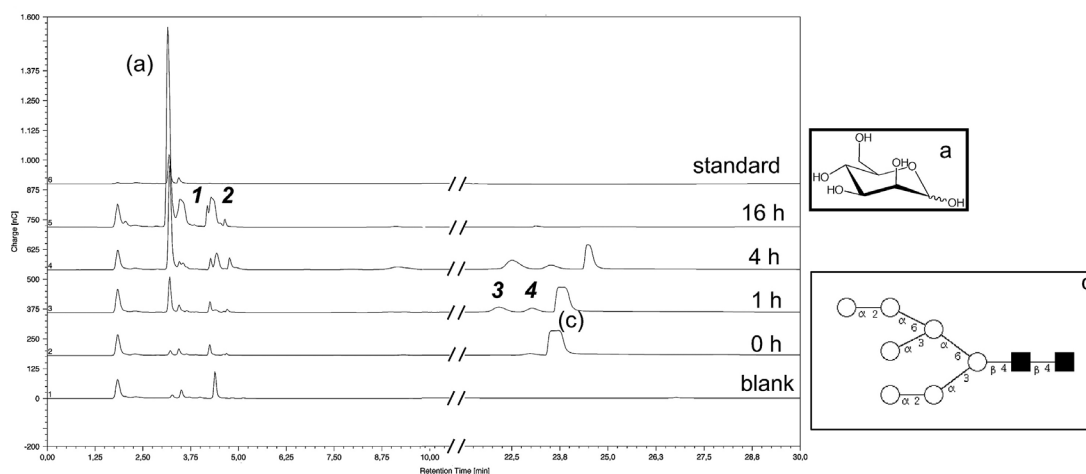
To understand the possible role *in vivo* of Ss $\alpha$ -man, we analysed the ability of the enzyme to hydrolyse mannosylated glycans found in N-glycosylated proteins.

The analysis by HPAE–PAD of the reaction mixtures containing Man<sub>3</sub>GlcNAc<sub>2</sub> showed that, already after 1 h of incubation, free mannose (a) and two well resolved peaks (1, and 2) with retention times lower than that of the substrate (b) could be identified (Figure. 2.6). After 16 h of incubation, Man<sub>3</sub>GlcNAc<sub>2</sub> was completely hydrolysed and the three products accumulated in the reaction.



**Figure 2.6:** Chromatographic runs of the mixtures containing nSs $\alpha$ -man and Man<sub>3</sub>GlcNAc<sub>2</sub>. The incubation times are also indicated on each run. Mannose (a), Man<sub>3</sub>GlcNAc<sub>2</sub> (b), were used as markers, and the blank mixture contains the enzyme only. In the schematic representation of Man<sub>3</sub>GlcNAc<sub>2</sub> mannose and N-acetylglucosamine are reported as empty circles and filled squares, respectively, and the anomeric bonds are also indicated.

In the case of Man<sub>7</sub>GlcNAc<sub>2</sub> (Figure. 3.6), after 1 h of incubation, we observed two new peaks (3, and 4) at retention times of 22–23 min that, together with the substrate (c), progressively disappear with the formation of mannose (a) and of the reaction products **1**, and **2**. The limited scale of the reaction precluded the structural determination of the observed product, thus, without suitable markers it is difficult to predict the structure of compounds **1–4**. Presumably, one of the products (1 or 2) could be Man<sub>1</sub>GlcNAc<sub>2</sub> containing a  $\beta$ -anomeric bond that cannot be hydrolysed by Ss $\alpha$ -man. Instead, the retention times of **3** and **4**, which are lower than that of the substrate, might be oligosaccharides resulting from the partial demannosylation of Man<sub>7</sub>GlcNAc<sub>2</sub>. However these experiments show that nSs $\alpha$ -man was able to recognize as substrates two high-mannose oligosaccharides commonly found in N-glycosylated proteins.

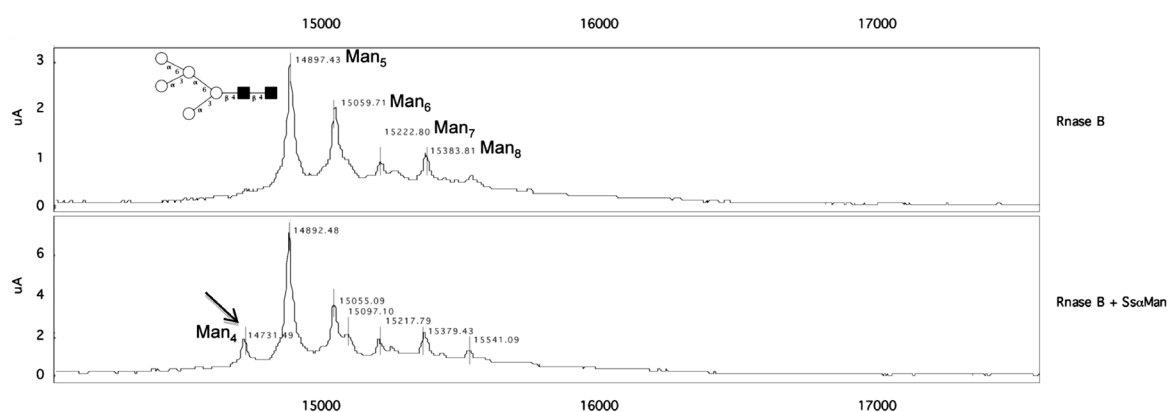


**Figure 3.6:** Chromatographic runs of the mixtures containing nSs $\alpha$ -man and Man<sub>7</sub>GlcNAc<sub>2</sub>. The incubation times are also indicated on each run. Mannose (a) and Man<sub>7</sub>GlcNAc<sub>2</sub>, (c) were used as markers, and the blank mixture contains the enzyme only. In the schematic representation of Man<sub>7</sub>GlcNAc<sub>2</sub> mannose and N-acetyl-glucosamine are reported as empty circles and filled squares, respectively, and the anomeric bonds are also indicated.

### Activity of SS $\alpha$ -Man on Ribonuclease B glycoprotein

The ability of the enzyme of hydrolysing high-mannose oligosaccharides prompted us to test nSs $\alpha$ -man on glycoproteins. Ribonuclease B (Rnase B) from bovine pancreas was chosen as a model substrate because it possesses a single N-glycosylation site with high-mannose oligosaccharides Man<sub>5</sub>–Man<sub>9</sub> (Tarelli et al. 2000).

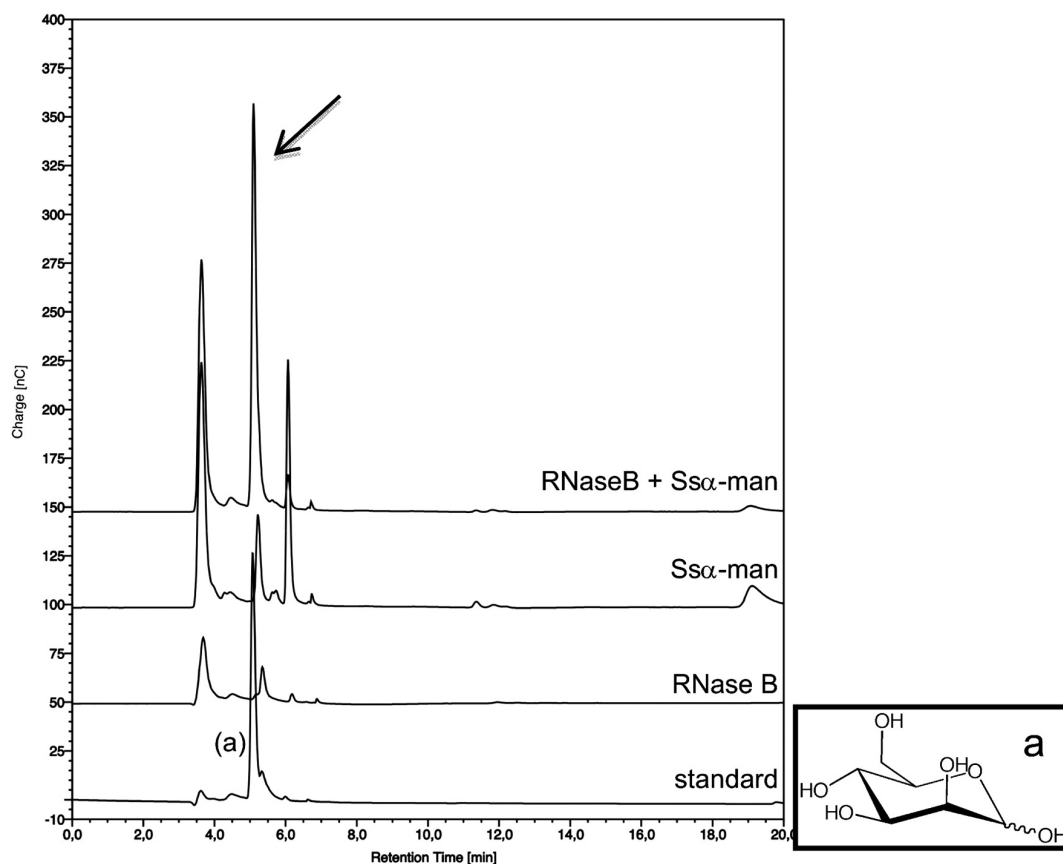
Remarkably, the analysis by SELDI-TOF MS of the reaction mixture, obtained by incubating nSs $\alpha$ -man and Rnase B at standard conditions, clearly shows the formation of a new peak at  $m/z$  14,731 next to the peak at  $m/z$  14,892 corresponding to the most abundant Rnase B/Man<sub>5</sub> glycoform (Figure 4.6).



**Figure 4.6:** SELDI-TOF mass spectrometer analysis of Rnase B incubated with Ss $\alpha$ -man. Reaction mixtures containing ribonuclease B (Rnase B) from bovine pancreas alone or together with nSs $\alpha$ -man are reported in the upper and lower part, respectively. In the upper part, the peaks corresponding to Man<sub>5</sub>–Man<sub>8</sub> are labelled correspondingly and the schematic structure of Man<sub>5</sub> is reported. In the lower part, the new peak at  $m/z$  14,731 is labelled with Man<sub>4</sub> and highlighted with an arrow

The difference in mass between the two peaks corresponds to that of a single mannose residue and fits well with those experimentally identified in the same SELDI-TOF MS spectrum for the different Rnase B glycoforms Man<sub>5</sub>/Man<sub>6</sub>/Man<sub>7</sub>/Man<sub>8</sub>/Man<sub>9</sub>. Therefore, this peak can be attributed to a novel Man<sub>4</sub> resulting from the single deglycosylation of Rnase B/Man<sub>5</sub> or, alternatively, from subsequent deglycosylations of Rnase B/Man<sub>6</sub>–Man<sub>9</sub>.





**Figure 5.6:** HPAE–PAD analysis of Rnase B incubated with Ssα-man. The samples analysed by HPAE–PAD were prepared at the same conditions described for those analysed by SELDI-TOF (see *Experimental procedures*). The peak observed in the sample containing both Rnase B and nSsα-man, showing the same retention time of d-mannose standard (a), is highlighted with an arrow.

Moreover, the analysis by HPAE–PAD of the same reaction mixture allowed us to detect free mannose after 16 h of incubation, confirming that Ssα-man is able to release mannose from Rnase B (Figure. 5.6). These data indicate that Ssα-man could be putatively involved in the turnover and/or the maturation of glycoprotein also *in vivo*.

## **Chapter IV**

### **Discussion**

In the laboratory where I have performed my thesis, a novel GH38  $\alpha$ -mannosidase from the thermoacidophilic Crenarchaeon *S. solfataricus* (Ss $\alpha$ -man), which is able to hydrolyse high-mannose oligosaccharides and mannosylated proteins, was characterized. This is the second archaeal  $\alpha$ -mannosidase characterized so far, and the first from Crenarchaea. The only other known archaeal  $\alpha$ -mannosidase (ManA) has been identified and characterized in the Euryarchaeon *P. torridus* (Angelov et al. 2006). Our study shows that the characteristics of the two enzymes are rather different. Both  $\alpha$ -mannosidases are intracellular, are expressed as single, leaderless, transcripts *in vivo*, are extremely selective for  $\alpha$ -D-mannosides, and are both inhibited by swainsonine and insensitive to 1-deoxymannojirimycin. Therefore, both can be grouped in class II  $\alpha$ -mannosidases. However, ManA and Ss $\alpha$ -man have only 25% sequence identity, diverging early in the GH38 phylogenetic tree, and, possibly, deriving from a lateral gene transfer.

The two enzymes show diverse oligomeric status and bind different metal cofactors: ManA is a dimer whose activity depends on  $\text{Cd}^{2+}$  while for Ss $\alpha$ -man, which is activated by  $\text{Zn}^{2+}$ , gel filtration experiments suggest that it is a trimer. This is not unusual among  $\alpha$ -mannosidases: the GH38 enzyme from *Bacillus*, 29% identical to Ss $\alpha$ -man, is trimeric (Nankai et al. 2002) as well as that from *Ginkgo biloba*, whose family is unknown (Woo et al. 2004). Several oligomeric status have been found, with dimeric (TmGH38, SpGH38, ManA), tetrameric (rat, pig, and human), and hexameric (yeast) enzymes (Yoshihisa et al. 1990; Yamashiro et al. 1997; Jin et al. 1999; Suits et al. 2010; Nakajima et al. 2003).

Also metal dependence in GH38 enzymes is rather diverse: cadmium activates the  $\alpha$ -mannosidase from *T. maritima* (Nakajima et al. 2003),  $\text{Co}^{2+}$  is the preferred cofactor for the  $\alpha$ -mannosidases from insect and *Bacillus* sp. (Nankai et al. 2002)(Kawar et al. 2001), while the activity of the enzymes from jack bean (Howard et al. 1997), dGMII, LAM, and SpGH38 is strictly dependent on  $\text{Zn}^{2+}$  (van den Elsen et al. 2001; Heikinheimo et al. 2003; Suits et al. 2010). In this last group of enzymes, this metal is particularly important being located in the  $\alpha/\beta$  portion in which forms an integral part of the -1 subsite of the catalytic site. Here, zinc coordinates the active site nucleophile and the  $\text{Man}_5$  in the  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  substrate (Numao et al. 2003; Shah et al. 2008). The importance of  $\text{Zn}^{2+}$  in catalysis and substrate binding is also confirmed in Ss $\alpha$ -man.

The recombinant enzyme that showed similar secondary structure, oligomeric status, and stability than the enzyme purified from *S. solfataricus*, however shows a 7.5-fold lower  $k_{\text{cat}}$  and presents only 2.3 zinc atoms per trimer, suggesting that the impaired activity results from the reduced efficiency in binding the cofactor.

Eukaryotic enzymes from GH38 are mainly involved in the maturation of glycoproteins during the transit in Golgi and ER or in their turnover in the lysosome. The function of cytosolic enzymes is less certain, but they are probably involved in protein recognition and signalling. The function of prokaryotic members has been less studied. The  $\alpha$ -mannosidase from *Bacillus* sp. participates in the degradation of xanthan, an exopolysaccharide of bacterial origin, allowing its exploitation as carbon source (Nankai et al. 2002). Apart from xanthan, not many exopolysaccharides contain  $\alpha$ -mannose; therefore, the possibility that the cytosolic Ss $\alpha$ -man could be involved *in vivo* in glycan degradation as a possible carbon source is questionable. In addition, 2- $\alpha$ -mannosyl-D-glycerate, a compatible solute very common in Euryarchaeota, is absent in *Sulfolobales*, thus, the involvement of Ss $\alpha$ -man in its

metabolism can be ruled out.

The  $\alpha$ -mannosidase from the human pathogen *M. tuberculosis* is involved in the catabolism of the cell wall glycoproteins and glycolipids whose terminal  $\alpha(1,2)$  mannose residues could function as virulence factors (Rivera-Marrero et al. 2001).  $\alpha$ -Mannose has been identified also in the cell envelope of other human parasites, but it is not a common component of bacterial glycoproteins therefore,  $\alpha$ -mannosidases from bacteria hosted in the human body might be involved in the catabolism of  $\alpha$ -mannosylated glycoconjugates, rather than in their maturation (Benz et al. 2002; Lillie et al. 2006). The function of the intracellular enzyme from another pathogen, *S. pyogenes*, is still unknown, however, it might be involved in the degradation of host glycans transported inside the bacterium (Suits et al. 2010).

Our study supports this conclusion. Ss $\alpha$ -Man purified from *S. solfataricus* extracts shows wide specificity for  $\alpha(1,2)$ ,  $\alpha(1,3)$ , and  $\alpha(1,6)$ -D-mannobiose substrates. In addition, analysis by HPAE–PAD showed that the enzyme is able to demannosylate Man<sub>3</sub>GlcNAc<sub>2</sub> and Man<sub>7</sub>GlcNAc<sub>2</sub> oligosaccharides commonly found in N-glycosylated proteins.

More interestingly, by SELDI-TOF mass spectrometry analysis, it was observed that Ss $\alpha$ -man removes mannose residues from the glycosidic moiety of the bovine pancreatic ribonuclease B, suggesting that it could process mannosylated proteins also *in vivo*. This is the first evidence that archaeal glycosidases are involved in the direct modification of glycoproteins (Cobucci-Ponzano et al. 2010).

GH38 collects a large variety of  $\alpha$ -mannosidases showing exquisite substrate specificity that has been described in detail giving clear indication on their function. In particular, the availability of the glycans that are their natural substrates *in vivo* allowed probing substrate specificities by 3D-structure and kinetic analyses. The characterization of the substrate specificity of Ss $\alpha$ -man is currently hampered by the limited information available on its natural substrates, precluding detailed functional studies. However, the abundant expression of Ss $\alpha$ -man *in vivo* and its ability in protein demannosylation makes this enzyme an interesting subject of study to further understand its function *in vivo*.

## References

Abu-Qarn, M., Eichler, J., and Sharon, N. (2008). Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Current Opinion in Structural Biology* 18, 544-550. Available at: <Go to ISI>://000260821300004.

Albert, M., Repetschnigg, W., Ortner, J., Gomes, J., Paul, B. J., Illaszewicz, C., Weber, H., Steiner, W., and Dax, K. (2000). Simultaneous detection of different glycosidase activities by F-19 NMR spectroscopy. *Carbohydrate Research* 327, 395-400. Available at: <Go to ISI>://000088527100005.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25, 3389-402. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=146917&tool=pmcentrez&rendertype=abstract>.

A. Amine, D. Moscone, R.A. Bernardo, E. Marconi, G. P. (2000). A new enzymatic spectrophotometric assay for the determination of lactulose in milk. *Analytica Chimica Acta* 406, 217-224. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0003267099007655>.

Amory, J. K. et al. (2007). Miglustat has no apparent effect on spermatogenesis in normal men. *Human reproduction* (Oxford, England) 22, 702-7. Available at: <http://humrep.oxfordjournals.org/cgi/content/abstract/22/3/702>.

Angelov, A., Putyrski, M., and Liebl, Wolfgang (2006). Molecular and biochemical characterization of alpha-glucosidase and alpha-mannosidase and their clustered genes from the thermoacidophilic archaeon *Picrophilus torridus*. *Journal of bacteriology* 188, 7123-31. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1636218&tool=pmcentrez&rendertype=abstract>.

Benz, I., and Schmidt, M. A. (2002). Never say never again: protein glycosylation in pathogenic bacteria. *Molecular Microbiology* 45, 267-276. Available at: <Go to ISI>://000176907100001.

Bommarius, A. S., and Riebel, B. R. (2004). *Biocatalysis: fundamentals and applications* (Wiley-VCH) Available at: <http://books.google.com/books?id=1rlBV6dhpkkC&pgis=1>.

Bone, Wilhelm, Walden, C. M., Fritsch, M., Voigtmann, U., Leifke, Eckhard, Gottwald, U., Boomkamp, S., Platt, F. M., and Spoel, A. C. van der (2007). The sensitivity of murine spermiogenesis to miglustat is a quantitative trait: a pharmacogenetic study. *Reproductive biology and endocrinology : RB&E* 5, 1. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1794412&tool=pmcentrez&rendertype=abstract>.

Boot, R. G., Verhoek, M., Donker-Koopman, W., Strijland, Anneke, Marle, J. van, Overkleeft, H. S., Wennekes, T., and Aerts, J. M. F. G. (2007b). Identification of the non-lysosomal glucosylceramidase as beta-glucosidase 2. *The Journal of biological chemistry* 282, 1305-12. Available at: <http://www.jbc.org/cgi/content/abstract/282/2/1305>.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72, 248-254. Available at: [http://dx.doi.org/10.1016/0003-2697\(76\)90527-3](http://dx.doi.org/10.1016/0003-2697(76)90527-3).

Bravman, T., Belakhov, V., Solomon, D., Shoham, G., Henrissat, B., Baasov, T., and Shoham, Y. (2003). Identification of the catalytic residues in family 52 glycoside hydrolase, a beta-xylosidase from *Geobacillus stearothermophilus* T-6. *Journal of Biological Chemistry* 278, 26742-26749. Available at: <Go to ISI>://000184155700058.

Burmeister, W. P., Cottaz, S., Rollin, P., Vasella, A., and Henrissat, B (2000). High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base. *The Journal of biological chemistry* 275, 39385-93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10978344>.

Campuzano, S., Pedrero, M., de Villena, F. Javier Manuel, and Pingarrón, J. (2004). A Lactulose Biosensor Based on Self-Assembled Monolayer Modified Electrodes. *Electroanalysis* 16, 1385-1392. Available at: <http://doi.wiley.com/10.1002/elan.200402983>.

Cantarel, B L, Coutinho, P M, Rancurel, C, Bernard, T, Lombard, V, and Henrissat, B (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* 37, D233-D238. Available at: <Go to ISI>://000261906200043.

Cobucci-Ponzano, B et al. (2009). beta-Glycosyl Azides as Substrates for alpha-Glycosynthases: Preparation of Efficient alpha-L-Fucosynthases. *Chemistry & Biology* 16, 1097-1108. Available at: <Go to ISI>://000271894000012.

Cobucci-Ponzano, B, Trincone, A, Giordano, A, Rossi, M, and Moracci, M (2003). Identification of the catalytic nucleophile of the family 29  $\alpha$ -L-fucosidase from *Sulfolobus solfataricus* via chemical rescue of an inactive mutant. *Biochemistry* 42, 9525-9531. Available at: <Go to ISI>://000184763000002.

Cobucci-Ponzano, Beatrice et al. (2010). A new archaeal beta-glycosidase from *Sulfolobus solfataricus*: seeding a novel retaining beta-glycan-specific glycoside hydrolase family along with the human non-lysosomal glucosylceramidase GBA2. *The Journal of biological chemistry* 285, 20691-703. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2898359&tool=pmcentrez&rendertype=abstract>.

Cobucci-Ponzano, Beatrice, Conte, Fiorella, Strazzulli, A., Capasso, C., Fiume, Immacolata, Pocsfalvi, G., Rossi, Mosè, and Moracci, Marco (2010). The molecular characterization of a novel GH38  $\alpha$ -mannosidase from the crenarchaeon *Sulfolobus solfataricus* revealed its ability in de-mannosylating glycoproteins. *Biochimie*. Available at: <http://dx.doi.org/10.1016/j.biochi.2010.07.016>.

Cobucci-Ponzano, Beatrice, Trincone, Antonio, Giordano, Assunta, Rossi, Mosè, and Moracci, Marco (2003). Identification of an archaeal  $\alpha$ -L-fucosidase encoded by an interrupted gene. Production of a functional enzyme by mutations mimicking programmed -1 frameshifting. *The Journal of biological chemistry* 278, 14622-31. Available at: <http://www.jbc.org/cgi/content/abstract/278/17/14622>.

Cobucci-Ponzano, Beatrice et al. (2010). A novel  $\alpha$ -D-galactosynthase from *Thermotoga maritima* converts  $\beta$ -D-galactopyranosyl azide to  $\alpha$ -galacto-oligosaccharides. *Glycobiology*. doi: 10.1093/glycob/cwq177.

Collins, T., De Vos, D., Hoyoux, A., Savvides, S. N., Gerday, C., Van Beeumen, J., et al. (2005). Study of the active site residues of a glycoside hydrolase family 8 xylanase. *Journal of molecular biology*, 354(2), 425-35. doi: 10.1016/j.jmb.2005.09.064.

Comfort, D. a, Bobrov, K. S., Ivanen, D. R., Shabalin, K. a, Harris, J. M., Kulminskaya, A. a, Brumer, H., and Kelly, Robert M (2007). Biochemical analysis of *Thermotoga maritima* GH36  $\alpha$ -galactosidase (TmGalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases. *Biochemistry* 46, 3319-30. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17323919>.

Côté, F., and Hahn, M. G. (1994). Oligosaccharins: structures and signal transduction. *Plant molecular biology* 26, 1379-411. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7858196>.

Darland, G., and Brock, T. (1971). *Bacillus acidocaldarius* sp. nov., an acidophilic thermophilic spore forming bacterium. *J Gen Microbiol* 69, 9-15.

Davies, B. (2000). Hand in Glove - Investigating Glycocode. *Chem. Ind*, 134-138. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20869790>.

Davies, G J, Wilson, K. S., and Henrissat, B (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochemical Journal* 321, 557-559. Available at: <Go to ISI>://A1997WE00400041.

Di Lauro, B., Strazzulli, A., Perugino, Giuseppe, La Cara, F., Bedini, Emiliano, Corsaro, Maria Michela, Rossi, Mosè, and Moracci, Marco (2008). Isolation and characterization of a new family 42 beta-galactosidase from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius*: identification of the active site residues. *Biochimica et biophysica acta* 1784, 292-301. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18068682>.

Dorling, P. R., Huxtable, C. R., and Colegate, S. M. (1980). Inhibition of lysosomal  $\alpha$ -mannosidase by swainsonine, an indolizidine alkaloid isolated from *Swainsona canescens*. *The Biochemical journal* 191, 649-51. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1162258&tool=pmcentrez&rendertype=abstract>.

Ducros, V. M. A., Tarling, C A, Zechel, D L, Brzozowski, A. M., Frandsen, T. P., Ossowski, I. von, Schulein, M., Withers, S G, and Davies, G J (2003). Anatomy of glycosynthesis: Structure and kinetics of the *Humicola insolens* Cel7B E197A and E197S glycosynthase mutants. *Chemistry & Biology* 10, 619-628. Available at: <Go to ISI>://000184378100008.

Elbein, A. D. (1991). Glycosidase inhibitors: inhibitors of N-linked oligosaccharide processing. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 5, 3055-63. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1743438>.

Elferink, M. G., Albers, S. V., Konings, W. N., and Driessen, A. J. (2001). Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. *Molecular microbiology* 39, 1494-503. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11260467>.

Elsen, J. M. H. van den, Kuntz, D A, and Rose, D R (2001). Structure of Golgi alpha-mannosidase II: a target for inhibition of growth and metastasis of cancer cells. *Embo Journal* 20, 3008-3017. Available at: <Go to ISI>://000169450000003.

Eng, C. M., Guffon, N., Wilcox, W. R., Germain, D. P., Lee, P., Waldek, S., Caplan, L., Linthorst, G. E., and Desnick, R. J. (2001). Safety and efficacy of recombinant human alpha-galactosidase A--replacement therapy in Fabry's disease. *The New England journal of medicine* 345, 9-16. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11439963>.

Fairweather, J. K., Faijes, M., Driguez, H., and Planas, A. (2002). Specificity studies of *Bacillus* 1,3-1,4-beta-glucanases and application to glycosynthase-catalyzed transglycosylation. *Chembiochem* 3, 866-873. Available at: <Go to ISI>://000177847600010.

Fridjonsson, O., and Mattes, R. (2001). Production of recombinant alpha-galactosidases in *Thermus thermophilus*. *Applied and environmental microbiology* 67, 4192-8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=93147&tool=pmcentrez&rendertype=abstract>.

Galili, U. (2001). The alpha-gal epitope (Gal alpha 1-3Gal beta 1-4GlcNAc-R) in xenotransplantation. *Biochimie* 83, 557-563. Available at: <Go to ISI>://000170846500002.

Ghazi, S., Rooke, J. A., and Galbraith, H. (2003). Improvement of the nutritive value of soybean meal by protease and alpha-galactosidase treatment in broiler cockerels and broiler chicks. *British poultry science* 44, 410-8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12964625>.

Goss, P. E., Reid, C. L., Bailey, D., and Dennis, J. W. (1997). Phase IB clinical trial of the oligosaccharide processing inhibitor swainsonine in patients with advanced malignancies. *Clinical cancer research : an official journal of the American Association for Cancer Research* 3, 1077-86. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9815786>.

Goto, K., Tanimoto, Y., Tamura, T., Mochida, K., Arai, D., Asahara, M., Suzuki, M., Tanaka, H., and Inagaki, K. (2002). Identification of thermoacidophilic bacteria and a new *Alicyclobacillus* genomic species isolated from acidic environments in Japan. *Extremophiles : life under extreme conditions* 6, 333-40. Available at: <http://www.springerlink.com/content/j894q3pbrmp25f0r/>.

Grabowski, G A, Gatt, S., and Horowitz, M. (1990). Acid beta-glucosidase: enzymology and molecular biology of Gaucher disease. *Critical reviews in biochemistry and molecular biology* 25, 385-414. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2127241>.

Grabowski, Gregory A (2008). Phenotype, diagnosis, and treatment of Gaucher's disease. *The Lancet* 372, 1263-1271. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19094956>.

Granovsky, M., Fata, J., Pawling, J., Muller, W. J., Khokha, R., and Dennis, J. W. (2000). Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nature medicine* 6, 306-12. Available at: <http://dx.doi.org/10.1038/73163>.

Györgydeák, Z., and Szilágyi, L. Einfache Synthesen der anomeren, an C-6 modifizierten Galacto- und Glucopyranosylazide. *Liebigs Annalen der Chemie* 1987, 235-241. Available at: <http://doi.wiley.com/10.1002/jlac.198719870314>.

Hancock, S. M., Vaughan, M. D., and Withers, Stephen G (2006). Engineering of glycosidases and glycosyltransferases. *Current opinion in chemical biology* 10, 509-19. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16905354>.

Hata, D. J., and Smith, D. S. (2004). Blood group B degrading activity of *Ruminococcus gnavus* alpha-galactosidase. Artificial cells, blood substitutes, and immobilization biotechnology 32, 263-74. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15274432>.

Heikinheimo, P. et al. (2003). The structure of bovine lysosomal alpha-mannosidase suggests a novel mechanism for low-pH activation. *Journal of molecular biology* 327, 631-44. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12634058>.

Helenius, A. (1994). Essay How N-linked Oligosaccharides Affect Glycoprotein Folding in the Endoplasmic Reticulum. *Molecular Biology of the Cell* 5, 253-265.

Henrissat, B, Callebaut, I., Fabrega, S., Lehn, P., Moron, J. P., and Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* 92, 7090-4. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=41477&tool=pmcentrez&rendertype=abstract>.



Hidaka, M., Fushinobu, S., Ohtsu, N., Motoshima, H., Matsuzawa, H., Shoun, H., and Wakagi, T. (2002). Trimeric Crystal Structure of the Glycoside Hydrolase Family 42  $\beta$ -Galactosidase from *Thermus thermophilus* A4 and the Structure of its Complex with Galactose. *Journal of Molecular Biology* 322, 79-91. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0022283602007465>.

Honda, Y., and Kitaoka, M. (2006). The first glycosynthase derived from an inverting glycoside hydrolase. *Journal of Biological Chemistry* 281, 1426-1431. Available at: <Go to ISI>://000234652000019.

Howard, S., Braun, C., McCarter, J., Moremen, K W, Liao, Y. F., and Withers, S G (1997). Human lysosomal and jack bean alpha-mannosidases are retaining glycosidases. *Biochemical and Biophysical Research Communications* 238, 896-898. Available at: <Go to ISI>://A1997YA66800040.

Hrmova, M., Imai, T., Rutten, S. J., Fairweather, J. K., Pelosi, L., Bulone, V., Driguez, H., and Fincher, G. B. (2002). Mutated barley (1,3)-beta-D-glucan endohydrolases synthesize crystalline (1,3)-beta-D-glucans. *Journal of Biological Chemistry* 277, 30102-30111. Available at: <Go to ISI>://000177509300093.

Huang, Y., Krauss, G., Cottaz, S., Driguez, H., and Lipps, G. (2005). A highly acid-stable and thermostable endo-beta-glucanase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Biochem. J.* Available at: <http://www.biochemj.org/bj/385/0581/bj3850581.htm>.

Jin, Y. Z., Dacheux, F., Dacheux, J. L., Bannai, S., Sugita, Y., and Okamura, N. (1999). Purification and properties of major alpha-D-mannosidase in the luminal fluid of porcine epididymis. *Biochimica et biophysica acta* 1432, 382-92. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10407159>.

Kawar, Z., Karaveg, K., Moremen, K W, and Jarvis, D. L. (2001). Insect cells encode a class II alpha-mannosidase with unique properties. *The Journal of biological chemistry* 276, 16335-40. Available at: <http://www.jbc.org/cgi/content/abstract/276/19/16335>.

Kiser, G. L., Gentzsch, M., Kloser, A. K., Balzi, E., Wolf, D. H., Goffeau, A., and Riordan, J. R. (2001). Expression and degradation of the cystic fibrosis transmembrane conductance regulator in *Saccharomyces cerevisiae*. *Archives of biochemistry and biophysics* 390, 195-205. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11396922>.

Knegtel, R. M. K., Strokopytov, B., Penninga, D., Faber, O. G., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L., and Dijkstra, B. W. (1995). Crystallographic Studies of the Interaction of Cyclodextrin Glycosyltransferase from *Bacillus Circulans* Strain-251 with Natural Substrates and Products. *Journal of Biological Chemistry* 270, 29256-29264. Available at: <Go to ISI>://A1995TJ22700037.

Koshland, D. E. (1953). STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. *Biological Reviews* 28, 416-436. Available at: <http://dx.doi.org/10.1111/j.1469-185X.1953.tb01386.x> DO - 10.1111/j.1469-185X.1953.tb01386.x.

Kosugi, A., Murashima, K., and Doi, R. H. (2002). Characterization of two noncellulosomal subunits, ArfA and BgaA, from *Clostridium cellulovorans* that cooperate with the cellulosome in plant cell wall degradation. *Journal of bacteriology* 184, 6859-65. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=135478&tool=pmcentrez&rendertype=abstract>.

Lal, A. (1996). Cloning, Expression, Purification, and Characterization of the Human Broad Specificity Lysosomal Acid alpha -Mannosidase. *Journal of Biological Chemistry* 271, 28348-28358. Available at: <http://www.jbc.org/cgi/content/abstract/271/45/28348>.

Leatherbarrow, R. J. (1992). GraFit.

Liebl, W., Wagner, B., and Schellhase, J. (1998). Properties of an alpha-galactosidase, and structure of its gene galA, within an alpha-and beta-galactoside utilization gene cluster of the hyperthermophilic bacterium *Thermotoga maritima*. *Systematic and applied microbiology* 21, 1-11. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9741105>.

Lillie, B. N., Hammermueller, J. D., MacInnes, J. I., Jacques, M., and Hayes, M. A. (2006). Porcine mannan-binding lectin A binds to *Actinobacillus suis* and *Haemophilus parasuis*. *Developmental and Comparative Immunology* 30, 954-965. Available at: <Go to ISI>://000240173900009.

Liu, Q. P. et al. (2007). Bacterial glycosidases for the production of universal red blood cells. *Nature biotechnology* 25, 454-64. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17401360>.

Macher, B. A., and Galili, U. (2008). The Gal alpha 1,3Gal beta 1,4GlcNAc-R (alpha-Gal) epitope: A carbohydrate of unique evolution and clinical relevance. *Biochimica Et Biophysica Acta-General Subjects* 1780, 75-88. Available at: <Go to ISI>://000253351600001.

Mackenzie, L. F., Wang, Q. P., Warren, R. A. J., and Withers, S G (1998). Glycosynthases: Mutant glycosidases for oligosaccharide synthesis. *Journal of the American Chemical Society* 120, 5583-5584. Available at: <Go to ISI>://000074213800030.

Malet, C., and Planas, A. (1998). From beta-glucanase to beta-glucansynthase: glycosyl transfer to alpha-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *Febs Letters* 440, 208-212. Available at: <Go to ISI>://000077348700044.

Malm, D., and Nilssen, O. (2008). Alpha-mannosidosis. *Orphanet Journal of Rare Diseases* 3, -. Available at: <Go to ISI>://000258375900001.

Marshall, R. O., and Kooi, E. R. (1957). Enzymatic conversion of D-glucose to D-fructose. *Science (New York, N.Y.)* 125, 648-9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/13421660>.

Martinez-Villaluenga, C., Cardelle-Cobas, A., Corzo, N., and Olano, A. (2008). Study of galactooligosaccharide composition in commercial fermented milks. *Journal of Food Composition and Analysis* 21, 540-544. Available at: <http://www.sciencedirect.com/science/article/B6WJH-4SPC0MX-3/2/d5f8a77fd0e587af3bbb883f6f23ac9d>.

Mast, S., and Moremen, K. (2006). Family 47  $\alpha$ -Mannosidases in N-Glycan Processing. *Methods in Enzymology* 415, 31-46. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17116466>.

Matern, H., Boermans, H., Lottspeich, F., and Matern, S. (2001). Molecular cloning and expression of human bile acid beta-glucosidase. *The Journal of biological chemistry* 276, 37929-33. Available at: <http://www.jbc.org/cgi/content/abstract/276/41/37929>.

Maurelli, L., Giovane, A., Esposito, A., Moracci, M., Fiume, I., Rossi, M., and Morana, A. (2008). Evidence that the xylanase activity from *Sulfolobus solfataricus* O alpha is encoded by the endoglucanase precursor gene (sso1354) and characterization of the associated cellulase activity. *Extremophiles* 12, 689-700. Available at: <Go to ISI>://000258758700008.

Mayer, C., Zechel, D L, Reid, S. P., Warren, R. A. J., and Withers, S G (2000). The E358S mutant of *Agrobacterium* sp beta-glucosidase is a greatly improved glycosynthase. *Febs Letters* 466, 40-44. Available at: <Go to ISI>://000085022000008.

McCarter, J. D., and Withers, S G (1994). Mechanisms of enzymatic glycoside hydrolysis. *Current opinion in structural biology* 4, 885-92. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7712292>.

McCutchen, C. M., Duffaud, G. D., Leduc, P., Petersen, A. R. H., Tayal, A., Khan, Saad A., and Kelly, Robert M. (2000). Characterization of extremely thermostable enzymatic breakers ( $\alpha$ -1,6-galactosidase and  $\beta$ -1,4-mannanase) from the hyperthermophilic bacterium *Thermotoga neapolitana* 5068 for hydrolysis of guar gum. *Biotechnology and Bioengineering* 52, 332-339. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18629900>.

Meier, H., and Reid, J. S. G. (1982). Reserve polysaccharides other than starch in higher plants. *Encyclopedia of plant physiology, N.S., vol. 13A: Plant carbohydrates I. Intracellular car.*

Mikami, B., Adachi, M., Kage, T., Sarikaya, E., Nanmori, T., Shinke, R., and Utsumi, S. (1999). Structure of raw starch-digesting *Bacillus cereus* beta-amylase complexed with maltose. *Biochemistry* 38, 7050-61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10353816>.

Montalto, M., Curigliano, V., Santoro, L., Vastola, M., Cammarota, G., Manna, R., Gasbarrini, A., and Gasbarrini, G. (2006). Management and treatment of lactose malabsorption. *World journal of gastroenterology : WJG* 12, 187-91. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16482616>.

Moracci, M., Trincone, A., Perugino, G., Ciaramella, M., and Rossi, M (1998). Restoration of the activity of active-site mutants of the hyperthermophilic beta-glycosidase from *Sulfolobus solfataricus*: Dependence of the mechanism on the action of external nucleophiles. *Biochemistry* 37, 17262-17270. Available at: <Go to ISI>://000077555800019.

Moremen, Kelley W (2002). Golgi alpha-mannosidase II deficiency in vertebrate systems: implications for asparagine-linked oligosaccharide processing in mammals. *Biochimica et biophysica acta* 1573, 225-35. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12417404>.

Møller, P. L., Jørgensen, F., Hansen, O. C., Madsen, S. M., and Stougaard, P. (2001). Intra- and extracellular beta-galactosidases from *Bifidobacterium bifidum* and *B. infantis*: molecular cloning, heterologous expression, and comparative characterization. *Applied and environmental microbiology* 67, 2276-83. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=92867&tool=pmcentrez&rendertype=abstract>.

Nakajima, M., Imamura, H., Shoun, H., and Wakagi, T. (2003). Unique metal dependency of cytosolic alpha-mannosidase from *Thermotoga maritima*, a hyperthermophilic bacterium. *Archives of biochemistry and biophysics* 415, 87-93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12801516>.

Nankai, H., Hashimoto, W., and Murata, K. (2002). Molecular identification of family 38 alpha-mannosidase of *Bacillus* sp strain GL1, responsible for complete depolymerization of xanthan. *Applied and Environmental Microbiology* 68, 2731-2736. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/122010015>.

Numao, S, He, S. M., Evjen, G, Howard, S., Tollersrud, O K, and Withers, S G (2000). Identification of Asp197 as the catalytic nucleophile in the family 38 alpha-mannosidase from bovine kidney lysosomes. *Febs Letters* 484, 175-178. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10800002>.

Numao, S, Kuntz, D A, Withers, S G, and Rose, D R (2003). Insights into the mechanism of *Drosophila melanogaster* Golgi alpha-mannosidase II through the structural analysis of covalent reaction intermediates. *Journal of Biological Chemistry* 278, 48074-48083. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12800087>.

Okuyama, M., Mori, H., Watanabe, K., Kimura, A., and Chiba, S. (2002). alpha-glucosidase mutant catalyzes "alpha-glycosynthase"-type reaction. *Bioscience Biotechnology and Biochemistry* 66, 928-933. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12500042>.

Overend, W. G. (1972). *The carbohydrates* 2nd ed. D. H. W. Pigman, ed. (New York and London : Academic Press).

Perugino, G, Cobucci-Ponzano, B, Rossi, M, and Moracci, M (2005). Recent advances in the oligosaccharide synthesis promoted by catalytically engineered glycosidases. *Advanced Synthesis & Catalysis* 347, 941-950. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1600006>.

Perugino, G, Trincone, A, Giordano, A, Oost, J. van der, Kaper, T., Rossi, M, and Moracci, M (2003). Activity of hyperthermophilic glycosynthases is significantly enhanced at acidic pH. *Biochemistry* 42, 8484-8493. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/124900012>.

Perugino, Giuseppe, Trincone, Antonio, Rossi, Mosé, and Moracci, Marco (2004). Oligosaccharide synthesis by glycosynthases. *Trends in biotechnology* 22, 31-37. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S016779903002865>.

Phenix, C. P., Rempel, B. P., Colobong, K., Doudet, D. J., Adam, M. J., Clarke, L. A., and Withers, Stephen G (2010). Imaging of enzyme replacement therapy using PET. *Proceedings of the National Academy of Sciences of the United States of America* 107, 10842-7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2890769&tool=pmcentrez&rendertype=abstract>.

Piirainen, L., Kekkonen, R. A., Kajander, K., Ahlroos, T., Tynkkynen, S., Nevala, R., and Korpela, R. (2008). In school-aged children a combination of galacto-oligosaccharides and *Lactobacillus* GG increases bifidobacteria more than *Lactobacillus* GG on its own. *Annals of nutrition & metabolism* 52, 204-8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18544974>.

Reiss, G., Heesen, S., Zimmerman, J., Robbins, P. W., and Aebi, M. (1996). the N-linked glycosylation pathway A ) B ). *Glycobiology* 6, 493-498.

Rempel, B. P., and Withers, Stephen G (2008). Covalent inhibitors of glycosidases and their applications in biochemistry and biology. *Glycobiology* 18, 570-586. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1800003>.

Rivera-Marrero, C. A., Ritzenthaler, J. D., Roman, J., and Moremen, K W (2001). Molecular cloning and expression of an alpha-mannosidase gene in *Mycobacterium tuberculosis*. *Microbial Pathogenesis* 30, 9-18. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1100002>.

Román-Leshkov, Y., Chheda, J. N., and Dumesic, J. a (2006). Phase modifiers promote efficient production of hydroxymethylfurfural from fructose. *Science (New York, N.Y.)* 312, 1933-7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16809536>.

Rydberg, E. H., Li, C. M., Maurus, R., Overall, C. M., Brayer, G. D., and Withers, S G (2002). Mechanistic analyses of catalysis in human pancreatic alpha-amylase: Detailed kinetic and structural studies of mutants of three conserved carboxylic acids. *Biochemistry* 41, 4492-4502. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12000036>.

Schwab, C., Sørensen, K. I., and Gänzle, M. G. (2010). Heterologous expression of glycoside hydrolase family 2 and 42 beta-galactosidases of lactic acid bacteria in *Lactococcus lactis*. *Systematic and applied microbiology*. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20822875>.

- Seeberger, P. H., and Werz, D. B. (2007). Synthesis and medical applications of oligosaccharides. *Nature* 446, 1046-51. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17460666>.
- Shah, N., Kuntz, Douglas A, and Rose, David R (2008a). Golgi alpha-mannosidase II cleaves two sugars sequentially in the same catalytic site. *Proc Natl Acad Sci U.S.A.* 105, 9570-9575. Available at: <http://www.pnas.org/content/105/28/9570.abstract>.
- Shaikh, F. A., Müllegger, J., He, S., and Withers, Stephen G (2007). Identification of the catalytic nucleophile in Family 42 beta-galactosidases by intermediate trapping and peptide mapping: YesZ from *Bacillus subtilis*. *FEBS letters* 581, 2441-6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17485082>.
- Shallom, D., Belakhov, V., Solomon, D., Gilead-Gropper, S., Baasov, T., Shoham, G., and Shoham, Y. (2002). The identification of the acid-base catalyst of alpha-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *FEBS Letters* 514, 163-167. Available at: <Go to ISI>://000175022400010.
- Shipkowski, S., and Brenchley, J. E. (2006). Bioinformatic, genetic, and biochemical evidence that some glycoside hydrolase family 42 beta-galactosidases are arabinogalactan type I oligomer hydrolases. *Applied and environmental microbiology* 72, 7730-8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1694227&tool=pmcentrez&rendertype=abstract>.
- Sinnott, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews* 90, 1171-1202. Available at: <http://dx.doi.org/10.1021/cr00105a006>.
- Siso, M. (1996). The biotechnological utilization of cheese whey: A review. *Bioresource Technology* 57, 1-11. Available at: [http://dx.doi.org/10.1016/0960-8524\(96\)00036-3](http://dx.doi.org/10.1016/0960-8524(96)00036-3).
- Sly, W. S., Quinton, B. A., McAlister, W. H., and Rimoin, D. L. (1973). Beta glucuronidase deficiency: report of clinical, radiologic, and biochemical features of a new mucopolysaccharidosis. *The Journal of pediatrics* 82, 249-57. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/4265197>.
- Spoel, A. C. van der, Jeyakumar, M., Butters, T. D., Charlton, H. M., Moore, H. D., Dwek, R. A., and Platt, F. M. (2002). Reversible infertility in male mice after oral administration of alkylated imino sugars: a nonhormonal approach to male contraception. *Proceedings of the National Academy of Sciences of the United States of America* 99, 17173-8. Available at: <http://www.pnas.org/cgi/content/abstract/99/26/17173>.
- Stam, M. R., Blanc, E., Coutinho, Pedro M, and Henrissat, Bernard (2005). Evolutionary and mechanistic relationships between glycosidases acting on alpha- and beta-bonds. *Carbohydrate research* 340, 2728-34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16226731>.
- Staudacher, E., Altmann, F., Wilson, I. B., and März, L. (1999). Fucose in N-glycans: from plant to man. *Biochimica et biophysica acta* 1473, 216-36. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10580141>.
- Suits, M. D. L., Zhu, Y., Taylor, E. J., Walton, J., Zechel, David L, Gilbert, H. J., and Davies, Gideon J (2010). Structure and kinetic investigation of *Streptococcus pyogenes* family GH38 alpha-mannosidase. *PloS one* 5, e9006. Available at: <http://dx.plos.org/10.1371/journal.pone.0009006>.
- Tanthanuch, W., Chantarangsee, M., Maneesan, J., and Ketudat-Cairns, J. (2008). Genomic and expression analysis of glycosyl hydrolase family 35 genes from rice (*Oryza sativa* L.). *BMC plant biology* 8, 84. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2531105&tool=pmcentrez&rendertype=abstract>.
- Tarelli, E., Byers, H. L., Wilson, M., Roberts, G., Homer, K. A., and Beighton, D. (2000). Detecting mannosidase activities using ribonuclease B and matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *Anal Biochem* 282, 165-172. Available at: <Go to ISI>://000088047200001.
- Tarling, C A, He, S. M., Sulzenbacher, G, Bignon, C, Bourne, Y, Henrissat, B, and Withers, S G (2003). Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from *Thermotoga maritima* through trapping of a covalent glycosyl-enzyme intermediate and mutagenesis. *Journal of Biological Chemistry* 278, 47394-47399. Available at: <Go to ISI>://000186731400008.
- Tarling, Chris A, He, S., Sulzenbacher, Gerlind, Bignon, Christophe, Bourne, Yves, Henrissat, Bernard, and Withers, Stephen G (2003). Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from *Thermotoga maritima* through trapping of a covalent glycosyl-enzyme intermediate and mutagenesis. *The Journal of biological chemistry* 278, 47394-9. Available at: <http://www.jbc.org/cgi/content/abstract/278/48/47394>.
- Tayal, A., Pai, V. B., and Khan, Saad A (1999). Rheology and Microstructural Changes during Enzymatic Degradation of a Guar-Borax Hydrogel. *Macromolecules* 32, 5567-5574. Available at: <http://dx.doi.org/10.1021/ma990167g>.

Terwisscha van Scheltinga, A. C., Armand, S., Kalk, K. H., Isogai, A., Henrissat, B., and Dijkstra, B. W. (1995). Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis. *Biochemistry* 34, 15619-23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7495789>.

Tolborg, J. F., Petersen, L., Jensen, K. J., Mayer, C., Jakeman, D. L., Warren, R. A. J., and Withers, S G (2002). Solid-phase oligosaccharide and glycopeptide synthesis using glycosynthases. *Journal of Organic Chemistry* 67, 4143-4149. Available at: <Go to ISI>://000176173700021.

Torarinsson, E., Klenk, H.-P., and Garrett, R. A. (2005). Divergent transcriptional and translational signals in Archaea. *Environmental microbiology* 7, 47-54. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15643935>.

Trincon, A., Perugino, G., Rossi, M., and Moracci, M (2000). A novel thermophilic glycosynthase that effects branching glycosylation. *Bioorganic & Medicinal Chemistry Letters* 10, 365-368. Available at: <Go to ISI>://000085539400014.

Umekawa, M., Huang, W., Li, B., Fujita, K., Ashida, H., Wang, L.-X., and Yamamoto, Kenji (2008). Mutants of *Mucor hiemalis* endo-beta-N-acetylglucosaminidase show enhanced transglycosylation and glycosynthase-like activities. *The Journal of biological chemistry* 283, 4469-79. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18096701>.

Van Laere, K. M., Abee, T., Schols, H. A., Beldman, G., and Voragen, A. G. (2000). Characterization of a novel beta-galactosidase from *Bifidobacterium adolescentis* DSM 20083 active towards transgalactooligosaccharides. *Applied and environmental microbiology* 66, 1379-84. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=91996&tool=pmcentrez&rendertype=abstract>.

Viladot, J. L., Canals, F., Batllori, X., and Planas, A. (2001). Long-lived glycosyl-enzyme intermediate mimic produced by formate re-activation of a mutant endoglucanase lacking its catalytic nucleophile. *Biochemical Journal* 355, 79-86. Available at: <Go to ISI>://000168086700010.

Viladot, J. L., Ramon, E. de, Durany, O., and Planas, A. (1998). Probing the mechanism of *Bacillus* 1,3-1,4-beta-D-glucan 4-glucanohydrolases by chemical rescue of inactive mutants at catalytically essential residues. *Biochemistry* 37, 11332-11342. Available at: <Go to ISI>://000075362900022.

Waclawovsky, A. J., Sato, P. M., Lembke, C. G., Moore, P. H., and Souza, G. M. (2010). Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant biotechnology journal* 8, 263-76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20388126>.

Wada, J. et al. (2008). 1,2-alpha-L-Fucosynthase: A glycosynthase derived from an inverting alpha-glycosidase with an unusual reaction mechanism. *Febs Letters* 582, 3739-3743. Available at: <Go to ISI>://000261234800006.

Watts, A. G., Opezzo, P., Withers, Stephen G, Alzari, P. M., and Buschiazzi, A. (2006). Structural and kinetic analysis of two covalent sialosyl-enzyme intermediates on *Trypanosoma rangeli* sialidase. *The Journal of biological chemistry* 281, 4149-55. Available at: <http://www.jbc.org/cgi/content/abstract/281/7/4149>.

Weely, S. van, Brandsma, M., Strijland, A., Tager, J. M., and Aerts, J. M. (1993). Demonstration of the existence of a second, non-lysosomal glucocerebrosidase that is not deficient in Gaucher disease. *Biochimica et biophysica acta* 1181, 55-62. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8457606>.

Williams, S. J., and Withers, S G (2000). Glycosyl fluorides in enzymatic reactions. *Carbohydrate research* 327, 27-46. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10968675>.

Withers, S G, and Aebersold, R. (1995). Approaches to Labeling and Identification of Active-Site Residues in Glycosidases. *Protein Science* 4, 361-372. Available at: <Go to ISI>://A1995QN56700002.

Withers, S G, Dombroski, D., Berven, L. A., Kilburn, D. G., Miller, R. C., Warren, R. A., and Gilkes, N. R. (1986). Direct <sup>1</sup>H n.m.r. determination of the stereochemical course of hydrolyses catalysed by glucanase components of the cellulase complex. *Biochemical and biophysical research communications* 139, 487-94. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3094517>.

Woo, K. K., Miyazaki, M., Hara, S., Kimura, M., and Kimura, Y. (2004). Purification and characterization of a co(II)-sensitive alpha-mannosidase from *Ginkgo biloba* seeds. *Bioscience, biotechnology, and biochemistry* 68, 2547-56. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15618626>.

Yamashiro, R., Itoh, H., Yamagishi, M., Natsuka, S., Mega, T., and Hase, S. (1997). Purification and Characterization of Neutral {alpha}-Mannosidase from Hen Oviduct: Studies on the Activation Mechanism of Co<sup>2+</sup>. *J. Biochem.* 122, 1174-1181. Available at: <http://jb.oxfordjournals.org/cgi/content/abstract/122/6/1174>.

Yang, F., Liu, Q., Bai, X., and Du, Y. (2010). Conversion of biomass into 5-Hydroxymethylfurfural Using Solid Acid Catalyst. *Bioresource Technology*. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0960852410016792>.

Yip, V. L. Y., Thompson, J., and Withers, Stephen G (2007). Mechanism of GlvA from *Bacillus subtilis*: a detailed kinetic analysis of a 6-phospho-alpha-glucosidase from glycoside hydrolase family 4. *Biochemistry* 46, 9840-52. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17676871>.

Yoshihisa, T., and Anraku, Y. (1990). A novel pathway of import of alpha-mannosidase, a marker enzyme of vacuolar membrane, in *Saccharomyces cerevisiae*. *The Journal of biological chemistry* 265, 22418-25. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2266133>.

Yoshimitsu, M. et al. (2004). Bioluminescent imaging of a marking transgene and correction of Fabry mice by neonatal injection of recombinant lentiviral vectors. *Proceedings of the National Academy of Sciences of the United States of America* 101, 16909-14. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=534735&tool=pmcentrez&rendertype=abstract>.

Zechel, D L, and Withers, S G (2001). Dissection of nucleophilic and acid-base catalysis in glycosidases. *Current Opinion in Chemical Biology* 5, 643-649. Available at: <Go to ISI>://000172563500004.

Zolghadr, B., Klingl, A., Koerdts, A., Driessen, A. J. M., Rachel, R., and Albers, S.-V. (2010). Appendage-mediated surface adherence of *Sulfolobus solfataricus*. *J. Bacteriol.* 192, 104-110. Available at: <http://jlb.asm.org/cgi/content/abstract/192/1/104>.

## List of Publications

- B. Cobucci-Ponzano, V. Aurilia, G. Riccio, B. Henrissat, P.M. Coutinho, A. Strazzulli, A. Padula, M.M. Corsaro, G. Pieretti, G. Pocsfalvi, I. Fiume, R. Cannio, M. Rossi, and M. Moracci, "A new archaeal beta-glycosidase from *Sulfolobus solfataricus*: seeding a novel retaining beta-glycan-specific glycoside hydrolase family along with the human non-lysosomal glucosylceramidase GBA2.," *The Journal of biological chemistry*, vol. 285, Jul. 2010, pp. 20691-703.
- B. Cobucci-Ponzano, F. Conte, A. Strazzulli, C. Capasso, I. Fiume, G. Pocsfalvi, M. Rossi, and M. Moracci, "The molecular characterization of a novel GH38 alpha-mannosidase from the crenarchaeon *Sulfolobus solfataricus* revealed its ability in de-mannosylating glycoproteins.," *Biochimie*, Aug. 2010.
- B. Cobucci-Ponzano, C. Zorzetti, A. Strazzulli, S. Carillo, E. Bedini, M.M. Corsaro, D.A. Comfort, R.M. Kelly, M. Rossi, and M. Moracci, "A novel  $\alpha$ -D-galactosynthase from *Thermotoga maritima* converts  $\beta$ -D-galactopyranosyl azide to  $\alpha$ -galacto-oligosaccharides," *Glycobiology*, 2010. (doi:10.1016/j.biochi.2010.07.016)
- B. Di Lauro, A. Strazzulli, G. Perugino, F. La Cara, E. Bedini, M.M. Corsaro, M. Rossi, and M. Moracci, "Isolation and characterization of a new family 42 beta-galactosidase from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius*: identification of the active site residues.," *Biochimica et biophysica acta*, vol. 1784, Feb. 2008, pp. 292-301.

## Meetings Communications

### CARBOHYDRATE ACTIVE ENZYMES IN GLYCOBIOLOGY: PHARMACEUTICAL TARGETS AND NEW TOOLS FOR THE SYNTHESIS OF INNOVATIVE CARBOHYDRATE-BASED DRUGS

Beatrice Cobucci-Ponzano, Andrea Strazzulli, Maria Carmina Ferrara, Carmela Zorzetti, Emiliano Bedini, Maria Michela Corsaro, Francesca Donaudy, Giancarlo Parenti, Mosè Rossi, and Marco Moracci. (DSV Conference 2010 - Rome, Oct. 11-12, 2010).

### THE CHARACTERIZATION OF A GH42 BETA-GALACTOSIDASE FROM THE THERMOPHILE *ALICYCLOBACILLUS ACIDOCALDARIUS* REVEALED A NOVEL CATALYTIC MACHINERY.

Strazzulli A, Perugino G, Cobucci-Ponzano B, Bedini E, Corsaro M.M, Fiume I, Withers S. G, Shaikh F, Rossi M, and Moracci M. (I Retreat Congiunto IGB – TIGEM - IBP; Roccaraso, 10-12 giugno 2009).

### CARBOHYDRATE ACTIVE ENZYMES FROM (HYPER)THERMOPHILES IN BIOTECHNOLOGY AND BIO-MEDICINE

Cobucci-Ponzano B, Strazzulli A, Michela Corsaro M, Bedini E, Aurilia V. 1, Pocsfalvi G. 1, Sulzenbacher G, Henrissat B, Rossi M, Moracci M. (I Retreat Congiunto IGB – TIGEM - IBP; Roccaraso, 10-12 giugno 2009).

**THE CHARACTERIZATION OF A THERMOPHILIC BETA-GLYCOSIDE HYDROLASE ENABLES TO CLASSIFY THE HUMAN BILE ACID BETA-GLUCOSIDASE IN A NEW GH FAMILY**

Cobucci-Ponzano B, Aurilia V, Riccio G, Strazzulli A, Corsaro M.M, Pieretti G, Pocsfalvi G, Fiume I, Henrissat B, Cannio R, Rossi M and Moracci M. (I Retreat Congiunto IGB – TIGEM - IBP; Roccaraso, 10-12 giugno 2009).

**NOVEL INSIGHT INTO THE REACTION MECHANISMS OF A GH42 BETA-GALACTOSIDASE**

A. Strazzulli, G. Perugino, B. Cobucci-Ponzano, E. Bedini, M.M Corsaro, I. Fiume, S.G. Withers, F. Shaikh, M. Rossi and M. Moracci. (8th Carbohydrate Bioengineering Meeting, 10-13 May 2009, Ischia, Naples).

**EFFECTS OF GLYCOSIDES ON GAG SYNTHESIS IN MUCOPOLY-SACCHARIDOSIS FIBROBLASTS**

F. Donaudy, A. Pignata, F. Fontana, B. Cobucci-Ponzano, A. Strazzulli, M. Moracci, and G. Parenti. (8th Carbohydrate Bioengineering Meeting, 10-13 May 2009, Ischia, Naples)

**IDENTIFICATION AND CHARACTERIZATION OF A BETA-GLYCOSIDE HYDROLASE FROM SULFOLOBUS SOLFATARICUS AS THE FIRST MEMBER OF A NEW GH FAMILY**

B. Cobucci-Ponzano, V. Aurilia, G. Riccio, A. Strazzulli, M.M. Corsaro, G. Pieretti, G. Pocsfalvi, B. Henrissat, R. Cannio, M. Rossi and M. Moracci. (8th Carbohydrate Bioengineering Meeting, 10-13 May 2009, Ischia, Naples)

**ENGINEERING GLYCOSIDASES FROM EXTREMOPHILES FOR THE SYNTHESIS OF OLIGOSACCHARIDES**

Beatrice Cobucci-Ponzano, Fiorella Conte, Andrea Strazzulli, Mosè Rossi, and Marco Moracci. (Extremophiles 2008 7-11 September 2008 Cape Town, South Africa).

**NOVEL MOLECULAR APPROACHES FOR THE MODIFICATION OF A GLYCOSIDASE INTO A GLYCOSYNTHASE.**

Andrea Strazzulli, Giuseppe Perugino, Beatrice Cobucci-Ponzano, Emiliano Bedini, Maria Michela Corsaro, Mosè Rossi and Marco Moracci. (3rd ERA-Chemistry "Flash" Conference - Carbohydrates at the interfaces of Biology, Medicine and Materials Science. 9th-13th March 2008, Killarney- Ireland).

**Foreign experience.**

Il Dott. Andrea Strazzulli è stato ospite dal 24 marzo 2010 al 1 luglio 2010 presso l'Istituto di Architecture et Fonction des Macromolécules Biologiques.-CNRS, Université de Provence, Université de la Méditerranée (Marseille), per la partecipazione al progetto "Nuove glicosidasi d'interesse terapeutico nella salute umana" Programma Galileo 2009/2010 finanziato dalla Università Italo-Francese (Coordinatore Dr. M. Moracci).





# Isolation and characterization of a new family 42 $\beta$ -galactosidase from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius*: Identification of the active site residues

Barbara Di Lauro <sup>a</sup>, Andrea Strazzulli <sup>a</sup>, Giuseppe Perugino <sup>a</sup>, Francesco La Cara <sup>a</sup>,  
Emiliano Bedini <sup>b</sup>, Maria Michela Corsaro <sup>b</sup>, Mosè Rossi <sup>a,c</sup>, Marco Moracci <sup>a,\*</sup>

<sup>a</sup> Institute of Protein Biochemistry, Consiglio Nazionale delle Ricerche, Via P. Castellino 111, 80131, Naples, Italy

<sup>b</sup> Dipartimento di Chimica Organica e Biochimica, Università di Napoli "Federico II",  
Complesso Universitario di Monte S. Angelo, Via Cinthia 4, 80126 Naples, Italy

<sup>c</sup> Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli "Federico II",  
Complesso Universitario di Monte S. Angelo, Via Cinthia 4, 80126 Naples, Italy

Received 20 June 2007; received in revised form 22 October 2007; accepted 29 October 2007

Available online 12 November 2007

## Abstract

The thermoacidophilic bacterium *Alicyclobacillus acidocaldarius* is a rich source of glycoside hydrolases enabling its growth on several di- and polysaccharides. We report here the purification and the characterization of a  $\beta$ -galactosidase from this source, the cloning of its gene, and the expression and the characterization of the recombinant enzyme (Aa $\beta$ -gal). The enzyme was purified 46-fold from *A. acidocaldarius* extracts; the gene for Aa $\beta$ -gal encoded a new member of the glycoside hydrolase family 42 (GH42) and it is flanked by a putative AraC/XylS regulator, however, the two genes were transcribed independently. The recombinant Aa $\beta$ -gal was characterized in detail revealing that it is optimally active and stable at 65 °C. Aa $\beta$ -gal is very specific for glycosides with an axial C4-OH at their non-reducing end, with  $k_{\text{cat}}/K_M$  values of 484, 186, and 332 s<sup>-1</sup> mM<sup>-1</sup> for 2-nitrophenyl- $\beta$ -D-galactoside, -fucoside, and 4-nitrophenyl- $\alpha$ -L-arabinoside, respectively. Finally, the characterization of the site-directed mutants Glu157Gly and Glu313Gly confirmed the latter as the nucleophile of the reaction and gave experimental evidence, for the first time in GH42, of the role of Glu157 as the acid/base of the catalyzed reaction.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** Extremophile; Thermophilic  $\beta$ -galactosidase; Reaction mechanism; Acid/base catalyst; Lactose hydrolysis; Lactulose; AraC/XylS

## 1. Introduction

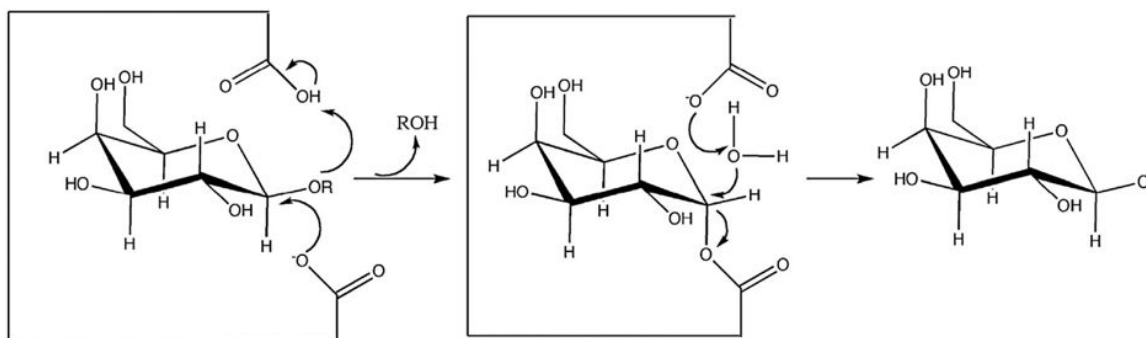
$\beta$ -Galactosidases (EC 3.2.1.23) are ubiquitous enzymes found in all three domains of life and are well known as gene reporters in molecular and cell biology and in the conversion of lactose in nutritional applications [1–3]. At present, an inspection of the carbohydrate-active enzyme classification CAZY (<http://www.cazy.org/>) shows that  $\beta$ -galactosidases belong to glycoside hy-

drolase (GH) families GH1, 2, 35, and 42; these families group enzymes which hydrolyse the substrate by following a *retaining* mechanism of hydrolysis in which the products have the same anomeric configuration as the substrate. Retaining glycoside hydrolases generally follow a double displacement mechanism involving a covalent glycosyl–enzyme intermediate (Scheme 1) [4]. In the first step of the reaction, one of the carboxyl groups functions as a general acid catalyst, protonating the glycosidic oxygen, while the other acts as a nucleophile. This leads to glycosidic bond cleavage, the departure of the leaving group, and the formation of a covalent glycosyl–enzyme intermediate. In the second step, the carboxylate deprotonates a water molecule, which, attacking the anomeric carbon, completes the hydrolysis producing the sugar with the same anomeric configuration of the

**Abbreviations:** GH, Glycoside hydrolase family; X-Gal, 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside; 2/4NP- $\beta$ -D-Gal, 2/4-nitrophenyl- $\beta$ -D-galactopyranoside; Fuc, fucoside; Ara, arabinoside; GST, Glutathione S-Transferase; UHT, Ultra High Temperature

\* Corresponding author. Tel.: +39 081 6132271; fax: +39 081 6132277.

E-mail address: [m.moracci@ibp.cnr.it](mailto:m.moracci@ibp.cnr.it) (M. Moracci).

Scheme 1. Reaction mechanism of retaining  $\beta$ -galactosidases.

substrate. As a consequence, the residues involved in catalysis are named the general acid/base and the nucleophile of the reaction, respectively.

Several 3D-structures are available for the enzymes of families GH1 and GH2 for which the catalytic residues have been experimentally determined. In contrast, the 3D-structure of only one enzyme is available for either the GH35 and the GH42 families [5,6] and, since both enzymes have been crystallized with the product galactose, the active site residues can be only inferred from these studies. Very recently, the group of Withers reported the unequivocal identification of the active site nucleophile of the reaction of the GH42  $\beta$ -galactosidase from *Bacillus subtilis* by using a mechanism-based inhibitor and subsequent peptide mapping [7]. However, for this family, the nature of the residue acting as the acid/base of the reaction, though proposed from the inspection of the 3D-structure of the enzyme from *Thermus thermophilus* A4 (Tt $\beta$ -gal) [5] and from the kinetic analysis of a mutant of the *B. subtilis*  $\beta$ -galactosidase [7], could not be assigned unequivocally.

On the basis of hydrophobic cluster analysis, GH42 enzymes are classified in the clan GH-A, which group 17 families showing the typical TIM barrel fold and following the *retaining* reaction mechanism [8]. To shed light on the catalytic machinery of GH42 enzymes we characterized in detail a  $\beta$ -galactosidase, belonging to this family, from the gram-positive, thermoacidophilic bacterium *Alicyclobacillus acidocaldarius*. This organism, which grows optimally at aerobic conditions, 60 °C and pH 3–4, can use as carbon energy source several di- and polysaccharides including melibiose, cellobiose, lactose, maltose, sucrose, trehalose, cellulose, xylan, starch, and glycogen [9,10], and, therefore, it has been recognized as a rich source of glycoside hydrolases [10–15]. Recently, we reported that *A. acidocaldarius* possesses two different  $\beta$ -gluco- and  $\beta$ -galactosidase activities and the functional cloning, the expression in *E. coli* of the gene encoding for a  $\beta$ -glucosidase (*glsB*), and the characterization of the enzyme, belonging to GH1, have been described in detail [14]. We show here the purification of the  $\beta$ -galactosidase enzyme, the cloning of the gene, and the detailed characterization of the recombinant enzyme. The gene encoding for this enzyme is flanked by a putative AraC/XylS regulator and the two genes are transcribed independently. Interestingly, the recombinant  $\beta$ -galactosidase was extremely specific for galacto-, fuco-,

and arabinoside substrates and it showed a catalytic efficiency on lactulose 7-fold higher than that on lactose. Finally, the unequivocal identification of the acid/base of the reaction by site-directed mutagenesis, the chemical rescue of the activity of the mutant and the analysis of the reaction products allowed the experimental identification of the residue acting as the acid/base of the reaction thereby clarifying, for the first time, the nature of the catalytic machinery of family GH42.

## 2. Materials and methods

### 2.1. Purification of Aa $\beta$ -gal from *A. acidocaldarius* extracts

The *A. acidocaldarius* strain ATCC 27009 growth conditions and the preparation of the crude extract have been described previously [14]. The extract was loaded onto a Q-Sepharose column for FPLC (Amersham Biotech, Sweden) equilibrated with 50 mM sodium phosphate buffer at pH 8.0 and eluted with a linear gradient of 0–0.5 M NaCl. The  $\beta$ -galactosidase activity was assayed in 50 mM sodium phosphate buffer at pH 6.5 on 5 mM 2NP- $\beta$ -D-Gal at 70 °C, and it was eluted at about 0.3 M NaCl. The active fractions were pooled and, after the addition of 1 M  $(\text{NH}_4)_2\text{SO}_4$ , they were applied onto a Hi-Load Phenyl-Sepharose reverse phase column for FPLC (Amersham Pharmacia) equilibrated with 50 mM sodium phosphate buffer at pH 7.0 and 1 M  $(\text{NH}_4)_2\text{SO}_4$ ; the enzyme bound to the column and it was eluted in water. The fractions containing  $\beta$ -galactosidase activity were pooled, concentrated, and equilibrated with 20 mM sodium phosphate buffer at pH 6.5. The sample was then applied to a Superose-O-6 gel filtration column for FPLC (Amersham Biotech) equilibrated with the same buffer; the active fractions were run on a SDS-PAGE to check their purity and then pooled and concentrated. The sample was stable for several months when stored in  $\text{NaN}_3$  0.02% (v/v) at 4 °C. Protein concentrations were determined with the method of Bradford [16], with bovine serum albumin as standard.

### 2.2. Cloning of the *lacB* gene

The preparation of the genomic library of *A. acidocaldarius* in the ZAP Express Pridigested Gigapack cloning kit (Stratagene, USA) and the functional screening on lactose and X-Gal were reported elsewhere [14]. The following degenerate oligonucleotide, deduced from the amino-terminal sequence of purified Aa $\beta$ -gal, was used for the screening of the library:

5'-GCI AAR CAY GCI CCI ATY TTY CCI AAY GTI CAR GGI TTY CTI  
CAY GGI GGI GAY TAY AAY-3'.

Triplets correspond to the amino acid sequence; I stands for inosine, R for A/G, and Y for C/T. The labelling of the oligonucleotide, the Southern blot analysis, and the screening of the library by plaque hybridization were performed by using standard molecular cloning techniques [17]. The phage library was used to infect

the *E. coli* strain Y1090; the 11 positive clones were isolated, singularly excised *in vivo* by following the indications of the manufacturer (Stratagene), and partially sequenced. The insert of the clone pBK-5.2Gal was sequenced on both strands.

### 2.3. Analysis of *lacB* expression in *A. acidocaldarius*

Total RNA from *A. acidocaldarius* cells were prepared as described previously [14] and the Northern blot was performed by following the protocol included in the Hybond-XL nylon membranes (Amersham Biotech). Each lane was loaded with 20 µg of total RNA. Each filter was hybridized under stringent conditions with specific DNA probes ( $4-8 \times 10^5$  c.p.m. mL<sup>-1</sup>). The probe for *lacB* was prepared by PCR from the region 2232–2795 bp (the first nucleotide of the fragment cloned from the *A. acidocaldarius* library was numbered as 1). The same filter used for the hybridization of *lacB* was stripped and re-hybridized with an *araC/xylS* specific probe (520–961 bp). Radioactivity was determined by autoradiography with a Storm PhosphorImager and visualized with the IQ-Mac software (Amersham Biotech).

### 2.4. Expression of the *lacB* gene and purification of the recombinant enzyme

The coding sequence of the *lacB* gene was amplified by PCR by using the *Pfu* DNA polymerase (Stratagene) and the phagemid pBK-5.2Gal as template. The oligonucleotides used for the amplification were the following (BamHI site is underlined):

5'Bam-Agal: 5'-CTCTGAGGATCCATGGCCAAGCACGCACCCATTTC-3'  
3'Bam-Agal: 5'-GAAACCGGATCCTCACCCATTTCGCGCCG-3'.

The amplified product was ligated to the expression vector pGEX-2TK (Amersham Biotech) as a fusion to the gene of the Glutathione *S*-transferase of *Schistosoma japonicum*. The recombinant plasmid obtained was named pGex-Aβgal and the insert was completely re-sequenced.

The cloning of the *lacB* gene in the expression vector pET29a was performed as described above by using the following amplification primer (NdeI site is underlined):

5'Nde-Agal: 5'-GCACCTATCATATGGCCAAGCACGCACCCATTTCC-3'

and the primer 3'Bam-Agal (see above), yielding the recombinant vector pETAβgal in which Aaβgal lacked any purification tag.

The recombinant β-galactosidase was expressed from the vector pGex-Aβgal in *E. coli* strain BL21RB791 cells and purified using the GST-tag and the thrombin cleavage on the matrix as described by the manufacturer (Amersham Biotech). After a single chromatographic step the enzyme was more than 95% pure by SDS-PAGE. The purification of the enzyme from pETAβgal/BL21 (DE3) *E. coli*, involved a thermoprecipitation at 60 °C and a Hi-Load Phenyl-Sepharose reverse phase column for FPLC, it was performed as previously described [14].

### 2.5. Characterization of Aaβgal

The standard assay for the β-galactosidase activity was performed in 50 mM sodium citrate buffer at pH 5.5 at 65 °C on the indicated substrates. Typically, in each assay we used 0.05–0.25 µg of Aaβgal in the final volume of 1.0 mL.

Kinetic constants of Aaβgal on aryl-glycosides were measured at standard conditions at 65 °C by using concentrations of substrate ranging between 0.05 and 30 mM. The εmM extinction coefficients for 2- and 4-nitrophenol under standard conditions and 65 °C were 1.1 and 3.7 mM<sup>-1</sup> cm<sup>-1</sup>, respectively. Kinetic constants on 2NP-β-D-Gal, at 65 °C, at different pHs were measured by using the following buffers at 50 mM concentrations: sodium citrate (pH 4.5–6.0), sodium acetate (pH 4.5–5.5), sodium phosphate (pH 5.5–8.0), and sodium borate (pH 7.5–8.5). The εmM extinction coefficients for 2-nitrophenol were accurately measured in each buffer, at 65 °C, and were in the range 0.2–1.8 mM<sup>-1</sup> cm<sup>-1</sup>.

Kinetic constants of lactose were measured as previously described by using 2.5 µg of enzyme and concentrations of substrate ranging between 2.5 and

700 mM [14]. Kinetic constants for change of lactulose were measured as above by using the Lactose/D-Galactose kit (Megazyme) by following the indications of the manufacturer. All kinetic data were calculated as the average of at least two experiments and were plotted and refined with the program GraFit [18].

The dependence of Aaβgal on metals was measured as previously described [14]. Molecular mass of denatured Aaβgal was determined on SDS-PAGE 10% in reducing conditions by using as molecular weight markers (Amersham Biotech, Sweden) phosphorylase b (97000), bovine serum albumin (66000), ovalbumine (45000), carbonic anhydrase (30000), trypsin inhibitor (20100), and α-lactalbumin (14400). Molecular mass of native Aaβgal was determined by gel filtration on a Superdex 200 HR 10/30 FPLC column (Amersham Biotech, Sweden); molecular weight markers were apoferritin (443000), β-glycosidase from *S. solfataricus* (240000) [19], bovine serum albumin (66000), and ovalbumine (43000).

Dependence on temperature was determined by assaying 0.05–5 µg of Aaβgal under standard conditions on 20 mM 4NP-β-D-Gal substrate in the indicated temperature range. The εmM extinction coefficients for 4-nitrophenol were accurately measured at each temperature and were in the range 1.3–4.3 mM<sup>-1</sup> cm<sup>-1</sup>. Thermal stability was tested by incubating pure Aaβgal (0.1 mg/mL) in 50 mM sodium phosphate buffer at pH 6.5 at the indicated temperatures. At intervals, aliquots were withdrawn from the mixture, transferred onto ice for 5 min, and assayed under standard conditions at 65 °C on 20 mM 2NP-β-D-Gal substrate. All the experiments were performed in triplicate.

### 2.6. Site-directed mutagenesis and characterization of the mutants

The mutants Glu157Gly and Glu313Gly were prepared by using the plasmid pGex-Aβgal, the following primers (mutations underlined):

E157G-mut: 5'-GTGATCGGCTGGCACGTGTGCAACGGGTACGG-CGGCG-3'  
E157G-rev: 5'-GTTTCGACACGTGGGAGCCGATCACGCCGGGATGATG-3'  
E313G-mut: 5'-CTCAAGAAGCCATTTCTGCTCATGGGGTCCACGCC-GAG-3'  
E313G-rev: 5'-CATGAGCAGAAATGGCTTCTTGAGGATGGCGCGG-3'

and the GeneTailor Site-Directed Mutagenesis System kit (Invitrogen). Mutations were identified by direct sequencing of the clones obtained and the genes encoding for the two mutants were completely resequenced to exclude the presence of accidental mutations. The mutant proteins were expressed and purified by exploiting the GST system described above.

The chemical rescue of the enzymatic activity of the mutants was performed as described in the text by using the 2NP-β-D-Gal substrate at the concentrations indicated.

### 2.7. Isolation and characterization of the products obtained by chemical rescue

To characterize the reaction products of the mutant Glu157Gly, 0.2 mg of enzyme were incubated at 65 °C for 16 h in 50 mM sodium citrate buffer at pH 4.6, sodium azide 0.2 M and 60 mg of 2NP-β-D-Gal in a total volume of 0.5 mL.

Isolation of β-D-galactosyl azide from the enzymatic reaction mixture was achieved by reverse phase column chromatography (Jupiter Proteo 90A, Phenomenex, 4 µm, 250 × 10 mm) on an Agilent HPLC instrument 1100 series, using H<sub>2</sub>O/CH<sub>3</sub>OH 9/1 as eluant. Positive-ions reflectron time-of-flight mass spectrum (MALDI-TOF-MS) was acquired on a Voyager DE-PRO instrument (Applied Biosystems) equipped with a delayed extraction ion source. Ion acceleration voltage was 20 kV, grid voltage was 14 kV, mirror voltage ratio 1.12 and delay time 100 ns. The sample was irradiated at a frequency of 5 Hz by 337-nm photons from a pulsed nitrogen laser. Mass calibration was obtained with a maltotrioligosaccharide mixture from corn syrup (Sigma). A solution of 2,5-dihydroxybenzoic acid in 20% CH<sub>3</sub>CN in water at a concentration of 25 mg/mL was used as the MALDI matrix. One microliter of matrix solution and 1 µL of the sample was premixed and then deposited on the target. The droplet was allowed to dry at ambient temperature. The spectrum was calibrated and processed under computer control using the Applied Biosystems Data Explorer software.

Table 1  
Purification of a  $\beta$ -galactosidase activity from *A. acidocaldarius* extracts<sup>1</sup>

Purification step	Total proteins (mg)	Total enzymatic activity (Units) <sup>2</sup>	Specific activity (U mg <sup>-1</sup> )	Purification folds	Yield (%)
Crude extract	560	7275	13	1	100
Ion-exchange chromatography	104	4996	48	3.7	69
Reverse phase chromatography	15	4605	307	24	63
Gel filtration	5	2960	592	46	41

<sup>1</sup>From 20 L culture and about 8 g of wet cell pellet.

<sup>2</sup>Assayed in 50 mM sodium phosphate buffer at pH 6.5 on 5 mM 2NP- $\beta$ -D-Gal at 70 °C.

NMR experiments (COSY and <sup>1</sup>H, <sup>13</sup>C HSQC) were recorded on a Bruker DRX400 Avance spectrometer using a 5 mm multinuclear inverse Z-gradient probe. <sup>13</sup>C and <sup>1</sup>H chemical shifts were measured in D<sub>2</sub>O using Acetone ( $\delta$  31.4 and  $\delta$  2225 for carbon and proton, respectively) as internal standard. <sup>1</sup>H (<sup>13</sup>C) NMR signal assignments: 4.60 H1 (<sup>3</sup>J<sub>H1,H2</sub>=8.5 Hz) (91.6 C1); 3.45 H2 (71.6 C2); 3.62 H3 (73.8 C3); 3.90 H4 (69.6 C4); 3.72 H5 (78.3 C5); 3.71 H6<sub>a,b</sub> (62.0 C6).

### 3. Results and discussion

#### 3.1. Identification and purification of a $\beta$ -galactosidase activity in *A. acidocaldarius*

We reported previously that, surprisingly, the functional screening on lactose or X-Gal of a genomic library of *A. acidocaldarius* produced only clones encoding for the  $\beta$ -glucosidase enzyme [14]. Therefore, we decided to purify the  $\beta$ -galactosidase from *A. acidocaldarius* and to clone the coding gene by 'reverse-genetics'.

The purification procedure is summarized in Table 1, showing a final yield of 41%, 46-fold purification, and a specific activity of 592 U mg<sup>-1</sup> in 50 mM sodium phosphate buffer at pH 6.5 on 5 mM 2NP- $\beta$ -D-Gal. Purified samples analyzed by SDS-PAGE were about 60% pure. The activity staining indicated that the purified sample had a migration similar to the  $\beta$ -galactosidase activity identified in the crude extract (Fig. 1A); instead, the  $\beta$ -glucosidase previously described migrated significantly slower [14]. The purified enzyme was named Aa $\beta$ -gal.

#### 3.2. Cloning of the gene encoding for the $\beta$ -galactosidase

The Edman degradation of Aa $\beta$ -gal produced the sequence AKHAFIPNVQGFLHGGDYN as the first 20 amino acids at the amino terminus. From this sequence we deduced a 60-mer degenerated oligonucleotide that, once tested by Southern blot of the *A. acidocaldarius* genomic DNA, showed a single signal and did not hybridize to the *glyB* gene (not shown); thus, the probe was specific and suitable for the following screening.

We isolated 11 clones out of 10<sup>5</sup> plaques screened; all the clones were ebscised and the corresponding phagemid DNA was purified. Among the clones tested we further analyzed the phagemid pBK-5.2Gal showing the largest insert of 3596 bp. This fragment revealed the presence on the same strand of two

ORFs of 966 bp and 2067 bp encoding for polypeptides of 322 and 688 amino acids, respectively. We deposited the entire *A. acidocaldarius* genomic fragment into GenBank with the accession number DQ092440.

The protein encoded by the longer ORF, showing a calculated molecular weight of 77,737 Da, displayed exactly the same amino-terminal sequence of the  $\beta$ -galactosidase from *A. acidocaldarius* except for a Val residue that was encoded by the first GTG codon. The absence of this amino acid in the functional enzyme suggested that it was removed after translation. The BLASTP (<http://www.ncbi.nlm.nih.gov/BLAST/>) analysis revealed that this ORF had high similarity to enzymes belonging to GH42; the highest identities were found with the  $\beta$ -galactosidases from *B. licheniformis* (61%), *B. circulans* (60%) *Yersinia pestis* (58%), *Klebsiella pneumoniae* (56%), and *Erwinia carotorova* (56%). In contrast, it showed lower identity with the  $\beta$ -galactosidase from *T. thermophilus* (32%). These data demonstrated that the cloned ORF encodes for the  $\beta$ -galactosidase identified in *A. acidocaldarius* extract and we named this gene *lacB*. Recently, another  $\beta$ -galactosidase gene from *A. acidocaldarius* ATCC 27009 virtually identical to *lacB* (97%) was cloned and expressed in *Pichia pastoris* [20].

Interestingly, the amino acid sequence of the ORF upstream to *lacB* gene, encoding for a protein with a calculated molecular weight of 36,467 Da, is homologous to the bacterial *araC/xylS*

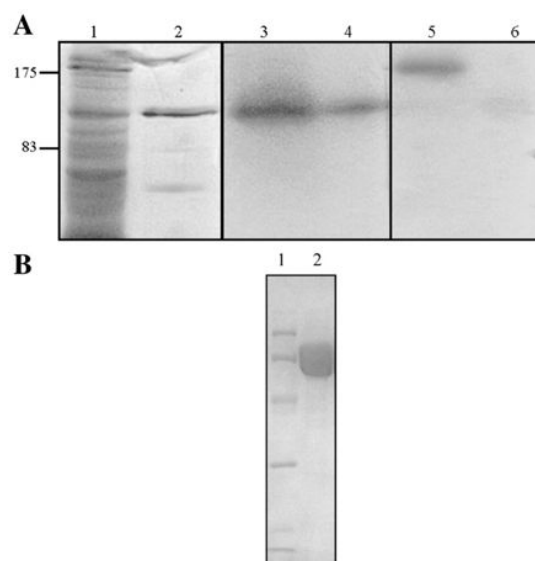


Fig. 1. Electrophoretic analysis of Aa $\beta$ -gal. (A) *A. acidocaldarius* extracts were analyzed by SDS-PAGE in lanes 1, 3, and 5 (85, 170, and 500  $\mu$ g, respectively); purified Aa $\beta$ -gal samples were loaded in lanes 2, 4, and 6 (5, 30, and 30  $\mu$ g, respectively). Lanes 1 and 2 were coomassie stained; lanes 3 and 4 were activity stained by incubation with X-Gal; lanes 5 and 6 were activity stained by incubation with X-Glc. The molecular weight markers are pre-stained  $\beta$ -galactosidase (175000) and paramyosin (83000). The samples were not heat denatured before loading; details of the activity staining are described elsewhere [14]. (B) SDS-PAGE of a purified sample of recombinant enzyme. Lane 1, molecular weight markers: phosphorylase b (97000), bovine serum albumin (66000), ovalbumine (45000), carbonic anhydrase (30000), trypsin inhibitor (20100), and  $\alpha$ -lactalbumin (14400). Lane 2, Aa $\beta$ -gal (28  $\mu$ g).



transcriptional regulators from *B. halodurans*, *Streptococcus suis* 89/1591, *Clostridium acetobutylicum*, and *B. subtilis* (47–49% similarity). The putative *araC/xylS* and *lacB* genes were encoded in the same translational frame and were separated by 75 bp. The cloned genomic fragment showed only a few nucleotides at the 5' non-coding region of *araC/xylS* gene precluding the identification of possible promoters upstream of this ORF. By using the program BROM (http://sun1.softberry.com/berry.phtml) we could not find any putative promoter in the region upstream of *lacB*; however, a purine-rich region, possibly acting as ribosome binding site, is located 9 bp from the first putative GTG codon of *lacB* (not shown). The lack of a clear promoter might explain why we could not functionally clone the *lacB* gene; presumably the expression level of this gene in *E. coli* was too low to observe the hydrolysis of lactose and X-Gal.

### 3.3. Transcriptional analysis of the *araC/xylS-lacB* gene cluster

To test if the identified genes were co-transcribed *in vivo*, we made a Northern blot of the total RNA extracted from *A. acidocaldarius* cells. Specific probes for *araC/xylS* and *lacB* genes revealed that the genes were transcribed independently (Fig. 2). The length of the *araC/xylS* transcript was about 960 bp, in agreement with the dimension of the gene (966 bp); instead, the *lacB* transcript was about 900 bp longer than expected. We could not identify other ORFs in the region downstream to the gene and the origin of this long transcript is obscure.

The presence of clustered AraC/XylS and  $\beta$ -galactosidase encoding genes is not uncommon, in *Staphylococcus xylosus* and *B. megaterium* DSM319 they are divergently arranged and the transcriptional regulator is involved in the regulation of  $\beta$ -galactosidase expression and in the utilization of lactose [21,22]. In *A. acidocaldarius* subspecies *rittmannii* it has been recently reported that lactose induced the  $\beta$ -galactosidase activity [15]. We could not find the binding site for activators of the AraC/XylS family upstream of the *lacB* gene in *A. acidocaldarius*; thus, how the expression of this gene is regulated *in vivo* will be matter of further investigations.

### 3.4. Expression and purification of the recombinant Aa $\beta$ -gal

The *lacB* gene was cloned and expressed in *E. coli* as a fusion with the *S. japonicum* Glutathione S-transferase (GST) gene and, without tags, in the vector pETA $\beta$ -gal. The former construct allowed us to purify the  $\beta$ -galactosidase from *A. acidocaldarius* in a single step of affinity chromatography, which routinely gave about 10 mg of pure protein from 1 L culture (Fig. 1B). After the removal of the GST portion, the enzyme showed a specific activity of 868 U mg<sup>-1</sup> in 50 mM sodium phosphate buffer at pH 6.5 on 5 mM 2-NP- $\beta$ -D-Gal at 70 °C, 1.5-fold higher than that of the  $\beta$ -galactosidase purified from *A. acidocaldarius* (see Table 1). Presumably, the higher specific activity originates from the higher purity of the recombinant enzyme. Increased expression of the *lacB* gene was obtained from the vector pETA $\beta$ -gal with final yields of 70 mg of pure enzyme per liter culture (not shown).

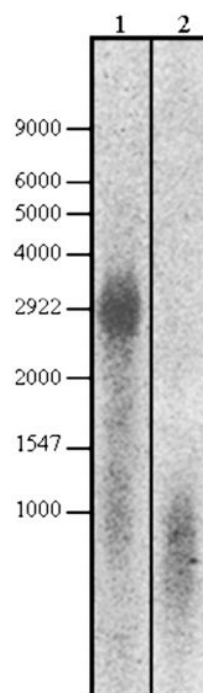


Fig. 2. Transcriptional analysis of the *lacB* locus. Northern blot of total RNA extracted from *A. acidocaldarius* cells. Each lane was loaded with 20  $\mu$ g of total RNA; lane 1, *lacB* gene, lane 2, *araC/xylS* gene.

The characteristics of the recombinant enzyme were compared to those of the  $\beta$ -galactosidase purified from *A. acidocaldarius* extracts, by means of the detailed enzymatic characterization of the two enzymes. No significant differences were observed, therefore, for brevity, we report here the data obtained from the recombinant enzyme, which was purified to homogeneity (Fig. 1B). The two enzymes showed identical molecular weights of 79,000  $\pm$  1000 and 162,000  $\pm$  1000 in denaturing and in native conditions, respectively, indicating that they are dimers in solution and that no significant post-translational modifications occurred in *A. acidocaldarius*. These molecular weights are very similar to those reported for the  $\beta$ -galactosidase from *A. acidocaldarius* subspecies *rittmannii* [15].

### 3.5. pH and temperature dependence

The pH vs log ( $k_{cat}/K_M$ ) behaviour on 2NP- $\beta$ -D-Gal substrate is reported in Fig. 3A. The enzymes produced a flattened bell-shaped curve at 65 °C (Fig. 3A). The optimal pH of the enzyme is close to neutrality and higher than the optimal growing pH of the acidophilic bacterium *A. acidocaldarius* (pH 4.0). This is not surprising, in fact, Aa $\beta$ -gal is an intracellular enzyme and its pH dependence confirmed its cytosolic location. At the optimal conditions the activity of Aa $\beta$ -gal was not affected by the presence of Mg<sup>2+</sup>, Mn<sup>2+</sup>, or Ca<sup>2+</sup> at 5 mM concentrations or by 12.5 mM EDTA, indicating that the enzyme catalyzes the hydrolysis of the substrate without the aid of divalent metal ion cofactors.

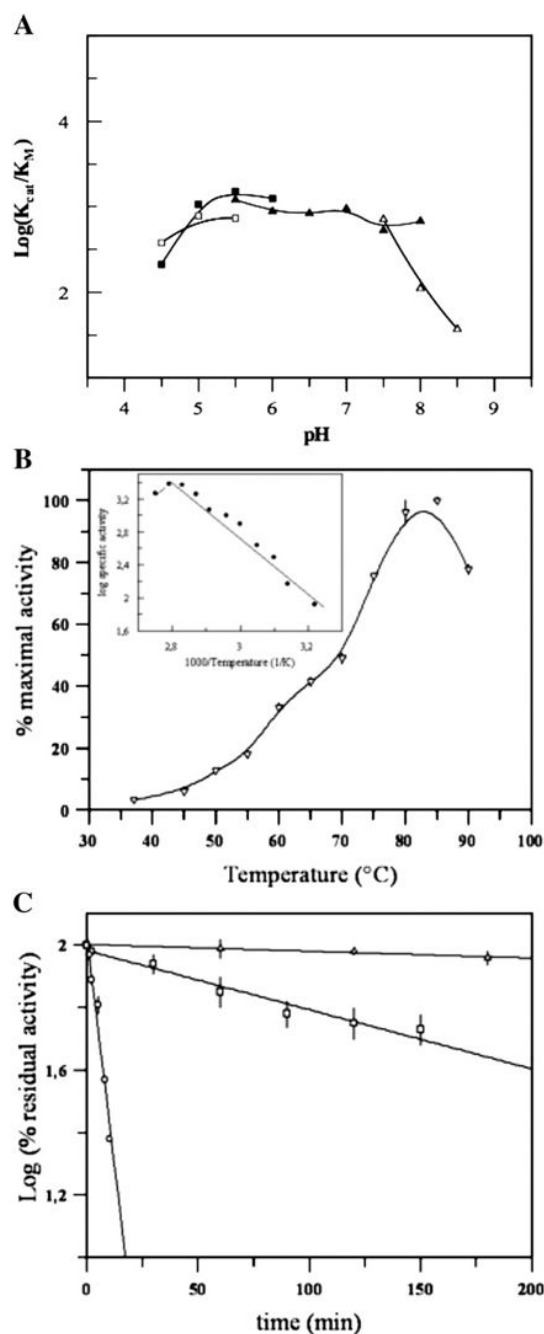


Fig. 3. pH Dependence, thermal activity and thermal stability of Aa $\beta$ -gal. (A) Catalytic efficiency of the enzyme on 2NP- $\beta$ -D-Gal is reported in the following buffers at 50 mM concentrations: sodium citrate (closed squares), sodium acetate (open squares), sodium phosphate (closed triangles), and sodium borate (open triangles). The spline fitting of the data points is reported to show the behaviour of the enzyme. (B) Thermal activity and derived Arrhenius plot (inset) of Aa $\beta$ -gal on 4NP- $\beta$ -D-Gal. (C) Thermal stability of Aa $\beta$ -gal at 65 °C (triangles), 75 °C (squares), and 80 °C (circles).

As expected for an enzyme from a thermophilic organism, the activity of Aa $\beta$ -gal is strictly dependent on temperature rising to a maximum at 85 °C (Fig. 3B). The linearity of the Arrhenius

plot demonstrated that the activation energy of the catalyzed reaction ( $E_a$ ) is not dependent on the temperature in the range 35–85 °C showing a value of  $71.67 \pm 3.97 \text{ kJ mol}^{-1}$ . The drop of activity at 90 °C presumably is due to the thermal denaturation of the enzyme at this temperature. Experiments of thermal stability supported this hypothesis demonstrating that Aa $\beta$ -gal is only barely stable at 80 °C (half-life 6 min) (Fig. 3C). However, the enzyme maintains 90% of the residual activity after 3 h at 65 °C. This is fully consistent with the optimal temperature of growth of *A. acidocaldarius* (60 °C). Again, the enzyme purified from *A. acidocaldarius* showed an almost identical behaviour.

### 3.6. Kinetic characterization

The substrate specificity of Aa $\beta$ -gal was analyzed by measuring the steady-state kinetic constants summarized in Table 2. The enzyme showed similar affinity ( $K_M$ ) for different arylglycosides while the catalytic efficiency ( $k_{cat}/K_M$ ) was in the order 2NP- $\beta$ -D-Gal > 4NP- $\alpha$ -L-Ara > 2NP- $\beta$ -D-Fuc. No activity was observed on glucosides and xylosides, which have their C4-OH group in the equatorial position, indicating that Aa $\beta$ -gal was extremely specific for the axial C4-OH group of the galacto-, fuco-, and arabinoside substrates. Among disaccharides, cellobiose, which has an equatorial C4-OH group, was not a substrate of the enzyme, while lactose and lactulose (4-*O*- $\beta$ -D-galactopyranosyl-D-fructofuranose) were hydrolysed efficiently.

The  $k_{cat}$  of the  $\beta$ -galactosidase purified from extracts of *A. acidocaldarius* subspecies *rittmannii*, assayed in sodium phosphate buffer is considerably lower, possibly because of the different buffer used, the lower purity of the sample if compared to our recombinant Aa $\beta$ -gal, or both [15]. In fact, also in our hands the  $k_{cat}$  values of the native Aa $\beta$ -gal were generally lower than those of the recombinant enzyme (not shown). However, the kinetic constants of the two enzymes followed a similar trend, giving interesting indications of the function of Aa $\beta$ -gal *in vivo*. The  $k_{cat}/K_M$  of lactose with Aa $\beta$ -gal is 10-fold lower than that of the  $\beta$ -glucosidase previously reported (namely  $32 \text{ s}^{-1} \text{ mM}^{-1}$ ) [14], thus, presumably, both enzymes support growth of *A. acidocaldarius* on lactose. However, the specificity of Aa $\beta$ -gal for arabinosides might indicate that it can eventually also allow growth on oligosaccharides from type I arabinogalactans. In fact, recently, it has been reported that several GH42  $\beta$ -galactosidases might hydrolyse oligosaccharides released from arabinogalactans by GH53 enzymes [23]. Therefore, it would be interesting to test if *A. acidocaldarius* can exploit these polysaccharides and if so, search in its extracts for arabinogalactanase activities.

Table 2  
Steady state kinetic constants of the recombinant Aa $\beta$ -gal

Substrate	$k_{cat}$ ( $\text{s}^{-1}$ )	$K_M$ (mM)	$K_{cat}/K_M$ ( $\text{s}^{-1} \text{ mM}^{-1}$ )
4NP- $\beta$ -d-Gal	$2988 \pm 78$	$4.9 \pm 0.4$	609
2NP- $\beta$ -d-Gal	$2657 \pm 86$	$5.5 \pm 0.6$	484
4NP- $\alpha$ -L-Ara	$1740 \pm 41$	$5.2 \pm 0.4$	332
2NP- $\beta$ -d-Fuc	$1345 \pm 43$	$7.3 \pm 0.6$	186
Lactose	$212 \pm 5$	$60.8 \pm 5.5$	3
Lactulose	$380 \pm 16$	$18.2 \pm 3.0$	21

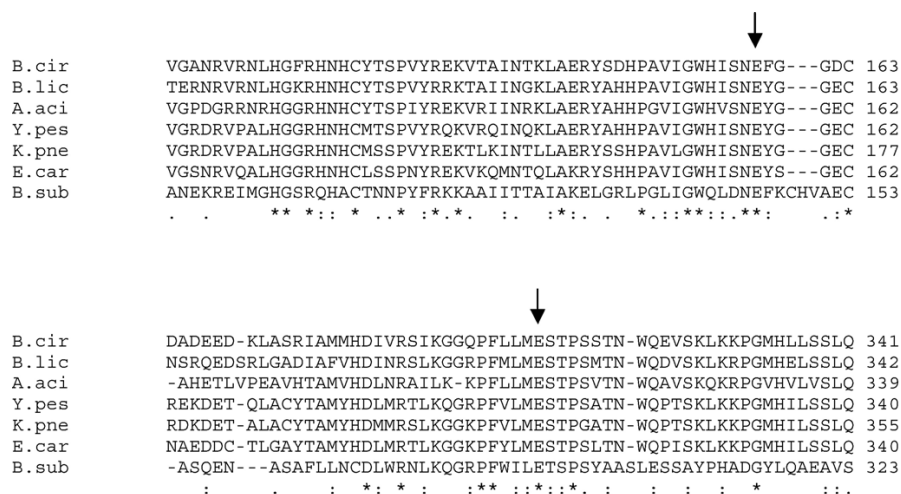
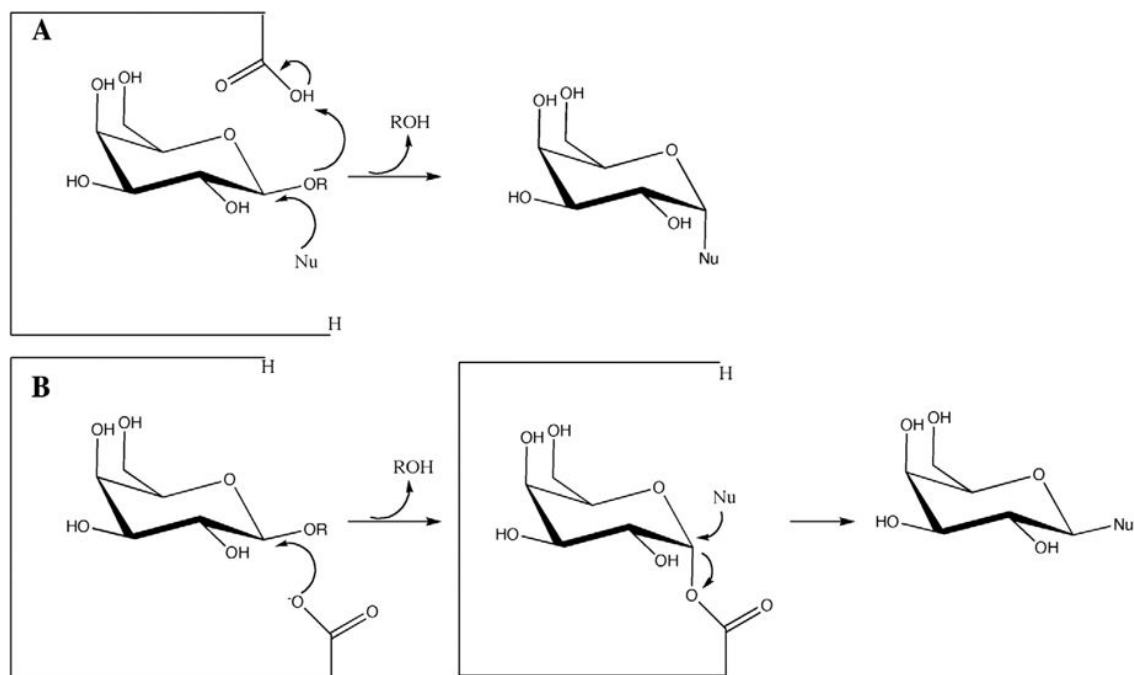


Fig. 4. Identification of the active site residues of Aa $\beta$ -gal. Multialignment of GH42  $\beta$ -galactosidases; the segment of the multiple sequence alignments in the region of the acid/base and the nucleophile are in the upper and the lower parts, respectively. The black arrows indicate the conserved catalytic glutamic acid residues; invariant and conserved amino acids are indicated by a "\*" and by ".", respectively. The multialignment was performed with the program ClustalW (<http://www.ebi.ac.uk/clustalw/>); the GenBank access numbers of the different genes used in the alignment were: *B. subtilis*, CAB12527.1; *B. circulans*, AAA22260; *B. licheniformis*, AAU22047; *Y. pestis*, AAS63701; *K. pneumoniae*, AAR53746; and *E. carotovora*, CAG76077.

It is worth noting the specificity of Aa $\beta$ -gal toward lactulose with a  $k_{\text{cat}}/K_M$  10-fold higher than that of the lactose. In the dairy industry lactulose is formed during milk heat treatment and it is the analytical index to distinguish UHT from sterilized milk by the International Dairy Federation and the European

Commission [see 24 and references therein]. Nowadays, the  $\beta$ -galactosidase from *Aspergillus oryzae* is preferred to the enzyme from *E. coli* for its higher relative activity on lactulose if compared to lactose [25], but enzymes showing higher optimal temperature and increased specificity for lactulose would be of



Scheme 2. Mechanisms of the chemical rescue of mutants in the nucleophile of the reaction (A) and the catalytic acid/base (B). Nu indicates the external nucleophile.

Table 3

Comparison of the wild type and mutants Aa $\beta$ -gal steady-state kinetic constants on 2NP- $\beta$ -D-Gal

Enzyme	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_{\text{M}}$ (mM)	$k_{\text{cat}}/K_{\text{M}}$ ( $\text{s}^{-1} \text{mM}^{-1}$ )
Wild type	$2657 \pm 86$	$5.5 \pm 0.6$	484
Glu313Gly	$0.21 \pm 0.04$	$4.8 \pm 0.4$	0.04
Glu157Gly	$0.32 \pm 0.02$	$0.11 \pm 0.04$	2.90

Kinetic constants were measured in 50 mM sodium citrate buffer at pH 5.5 at 65 °C.

great biotechnological potential. Therefore Aa $\beta$ -gal, which is thermally stable and expressed at high level, is a promising diagnostic tool for the determination of lactulose in milk.

### 3.7. Identification of the active site residues

The inspection of the 3D-structure of Tt $\beta$ -gal in complex with galactose suggested that the acid/base and the nucleophile of the reaction correspond to Glu141 and Glu312, respectively [5]. The function of the latter has recently been unequivocally confirmed on the  $\beta$ -galactosidase from *B. subtilis* for the residue Glu295 (*B. subtilis* numbering) while the acid/base of the same enzyme was tentatively assigned as Glu145 [7]. The multiple sequence alignment of the amino acid sequence of Aa $\beta$ -gal enzyme with those of other GH42 enzymes indicates that in the  $\beta$ -galactosidase from *A. acidocaldarius* these residues correspond to Glu157 and Glu313, which, as expected for catalytic residues, are invariant in family 42 (Fig. 4).

In an effort to experimentally prove the role played in catalysis by Aa $\beta$ -gal Glu157 and Glu313, we prepared the mutants Glu157Gly and Glu313Gly and we applied the chemical rescue strategy to analyze their kinetic behaviour (Scheme 2). Briefly, when the catalytic residues of glycoside hydrolases are replaced by non-nucleophilic amino acids the mutants are severely impaired in their activity. The addition to the reaction mixture of small external anions, such as azide or formate, leads to the reactivation, the so-called *chemical rescue* of the enzymatic activity, in the presence of an activated substrate. Furthermore, the analysis of the anomeric stereochemistry of the products obtained with the mutant reactivated with sodium azide allows the unequivocal identification of the function of the residue mutated [26]. In fact, in the case of the mutation of the nucleophile, the activity rescued with azide leads to products with the anomeric configuration *inverted* compared to that of the substrate: i.e. production of  $\alpha$ -glycosyl azide by mutated  $\beta$ -glycosidases (Scheme 2A). Instead, as shown in Scheme 2B, the reactivation of the mutant in which the general acid/base residue has been replaced yields product with *retained* anomeric configuration ( $\beta$ -glycosyl azide product from mutant  $\beta$ -glycosidases). The reason for this behaviour is that the azide ion replaces the nucleophile residue in the first step of the reaction or the general base in the second step, respectively, depending on the mutant analyzed (Scheme 2).

Glu157Gly and Glu313Gly mutants were expressed and purified as GST fusions by following the protocol described above for the wild type enzyme. As expected, the removal of Glu313 severely inactivated the enzyme and the  $k_{\text{cat}}/K_{\text{M}}$  for the

mutant Glu313Gly, assayed on 2NP- $\beta$ -D-Gal in 50 mM sodium citrate buffer at pH 5.5 at 65 °C was reduced by  $1.2 \times 10^4$ -fold compared to the wild type (Table 3). However, no reactivation of the Glu313Gly mutant was observed when assayed on the same substrate in the presence of concentrations of sodium azide ranging between 20 mM and 1 M. The absence of reactivation in sodium azide precluded the determination of the anomeric configuration of the reaction products. Presumably, high temperature and sodium azide together inactivated the mutant. In contrast, in the presence of sodium formate, we observed a progressive reactivation of the mutant resulting in a 15-fold increase of the enzymatic activity at a final concentration of 2 M of this nucleophile (Fig. 5). This level of reactivation is lower than that observed with other  $\beta$ -retaining enzymes, which sometimes even approach wild type levels [27], but it is comparable to that reported for  $\alpha$ -retaining enzymes [28–30]. We have reported previously that  $\beta$ -glycosidases mutated in their nucleophile and reactivated by sodium formate and activated substrates are able to promote the synthesis of oligosaccharides and were thereby named glycosynthases [19]. Unfortunately, the Aa $\beta$ -gal mutant Glu313Gly reactivated with this method did not act as a glycosynthase and we could not observe the formation of transgalactosylation products.

The turnover number of the Glu157Gly mutant on the 2NP- $\beta$ -D-Gal substrate was  $0.32 \pm 0.02 \text{ s}^{-1}$ , which is  $8.3 \times 10^3$ -fold lower than that of the wild type assayed under the same conditions (Table 3). More interestingly, the affinity of the mutant for this substrate was 50-fold higher than that of the wild type ( $K_{\text{M}}$  of  $0.11 \pm 0.04 \text{ mM}$  and  $5.5 \pm 0.6 \text{ mM}$ , for the mutant and the wild type respectively). This result, which, despite the different mutation, is similar to that observed on the Glu145Ala mutant of *B. subtilis*  $\beta$ -galactosidase on the 4NP- $\beta$ -D-Gal substrate, is an indication that Glu157 is the active site acid/base [7] since the reduction of the  $K_{\text{M}}$  results from the accumulation of the galactosyl–enzyme intermediate.

The comparison of the pH dependence of the Glu157Gly mutant with that of the wild type showed that the mutation flattened the typical bell-shaped curve of the wild type Aa $\beta$ -gal

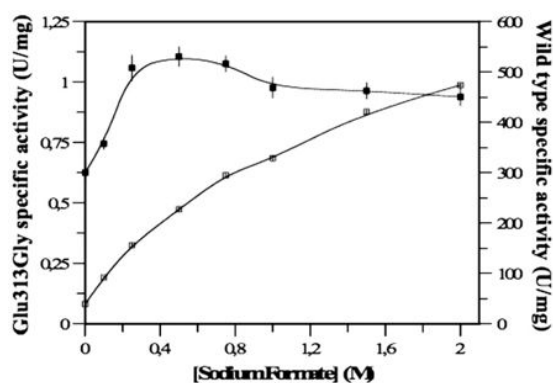


Fig. 5. Chemical rescue of the Glu313Gly mutant. The wild type (closed symbols) and mutant (open symbols) Aa $\beta$ -gal were assayed on 12 mM 2NP- $\beta$ -D-Gal in 50 mM sodium citrate buffer at pH 5.5 at 65 °C in the presence of sodium formate.



(not shown). This is another strong indication of the function of Glu157 as the catalytic acid/base residue; however, a more definitive test for having mutated the acid/base catalyst is the analysis of the chemical rescue in the presence of external nucleophiles. The activity of the Glu157Gly mutant was not recovered in the presence of sodium formate at concentrations ranging between 0.2 and 1 M even after prolonged incubation. Instead, the same mutant showed a limited rescue of activity (2-fold) on 12 mM 2NP- $\beta$ -D-Gal after 16 h of incubation at standard conditions in the presence of sodium azide 200 mM. Under these conditions, the mutant produced a compound that once isolated and characterized proved to be  $\beta$ -galactosyl azide. The identity of the product was deduced on the basis of its positive ion MALDI-TOF-MS spectrum, that showed a signal at  $m/z$  228.045 attributed to the pseudomolecular ion  $[(M+Na)^+]$  calculated accurate mass 228.059 Da of a hexosyl azide. The galactose spin system and the configuration of the anomeric carbon were deduced from NMR experiments. In particular the value of  $^3J_{H1,H2}$  of 8.5 Hz indicated a beta configuration, as also observed for  $\beta$ -L-fucosyl azide [29].

The retained configuration of the novel product  $\beta$ -galactosyl azide (Scheme 2B) and the observed increased rate of the reaction in the presence of sodium azide are the results of the more rapid reaction of azide with the galactosyl-enzyme intermediate if compared to water in the absence of the general base catalyst. These effects were not observed with the wild type Aa $\beta$ -gal, in which, presumably, the charge screening of the Glu157 precludes the access of the azide.

These results provide the direct evidence that the residue Glu157 is the acid/base of the reaction and allow us to assign experimentally for the first time this important catalytic function to a family GH42 enzyme. This finding can be easily extended to all the other members of this family.

## Acknowledgments

We thank Vito Carratore and Laura Camardella for the protein sequencing, and Giuseppe Ruggiero for the *A. acidocaldarius* growth. This work was partially supported by the 'MIUR-Decreto Direttoriale prot. n. 1105/2002' and by the Eurochem S.P.A. project S517-P "Biosensori innovativi che impiegano enzimi dotati di particolari proprietà isolati anche da micro-organismi termofili". We thank the TIGEM-IGB DNA sequencing core for the sequencing of the clones. The IBP-CNR belongs to the Centro Regionale di Competenza in Applicazioni Tecnologico-Industriali di Biomolecole e Biosistemi.

## References

- [1] V. Gekas, M. Lopez-Leiva, Hydrolysis of lactose: a literature review, *Process Biochem.* 20 (1985) 2–12.
- [2] N.S. Scrimshaw, E.B. Murray, The acceptability of milk and milk products in populations with a high prevalence of lactose intolerance, *Am. J. Clin. Nutr.* 48 (1988) 1079–1159.
- [3] M.I.G. Siso, The biotechnological utilization of cheese whey: a review, *Bioresour. Technol.* 57 (1996) 1–11.
- [4] D.E. Koshland Jr., Stereochemistry and the mechanism of enzymatic reactions, *Biol. Rev.* 28 (1953) 416–436.
- [5] M. Hidaka, S. Fushinobu, N. Ohtsu, H. Motoshima, H. Matsuzawa, H. Shoun, T. Wakagi, Trimeric crystal structure of the glycoside hydrolase family 42 beta-galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose, *J. Mol. Biol.* 322 (2002) 79–91.
- [6] A.L. Rojas, R.A. Nagem, K.N. Neustroev, M. Arand, M. Adamska, E.V. Eneyskaya, A.A. Kulminkaya, R.C. Garratt, A.M. Golubev, I. Polikarpov, Crystal structures of beta-galactosidase from *Penicillium* sp. and its complex with galactose, *J. Mol. Biol.* 343 (2004) 1281–1292.
- [7] F.A. Shaikh, J. Mullegger, S. He, S.G. Withers, Identification of the catalytic nucleophile in family 42 beta-galactosidases by intermediate trapping and peptide mapping: YesZ from *Bacillus subtilis*, *FEBS Lett.* 581 (2007) 2441–2446.
- [8] B. Henrissat, I. Callebaut, S. Fabrega, P. Lehn, J.P. Mornon, G. Davies, Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases, *Proc. Natl. Acad. Sci. U. S. A.* 92 (1995) 7090–7094.
- [9] K. Goto, Y. Tanimoto, T. Tamura, K. Mochida, D. Arai, M. Asahara, M. Suzuki, H. Tanaka, K. Inagaki, Identification of thermoacidophilic bacteria and a new *Alicyclobacillus* genomic species isolated from acidic environments in Japan, *Extremophiles* 6 (2002) 333–340.
- [10] K. Eckert, E. Schneider, A thermoacidophilic endoglucanase (CelB) from *Alicyclobacillus acidocaldarius* displays high sequence similarity to arabinofuranosidases belonging to family 51 of glycoside hydrolases, *Eur. J. Biochem.* 270 (2003) 3593–3602.
- [11] K. Eckert, F. Zielinski, L. Lo Leggio, E. Schneider, Gene cloning, sequencing, and characterization of a family 9 endoglucanase (CelA) with an unusual pattern of activity from the thermoacidophile *Alicyclobacillus acidocaldarius* ATCC27009, *Appl. Microbiol. Biotechnol.* 60 (2002) 428–436.
- [12] T.T. Koivula, H. Hemila, R. Pakkanen, M. Sibakov, I. Palva, Cloning and sequencing of a gene encoding acidophilic amylase from *Bacillus acidocaldarius*, *J. Gen. Microbiol.* 139 (1993) 2399–2407.
- [13] J. Matzke, A. Herrmann, E. Schneider, E.P. Bakker, Gene cloning, nucleotide sequence and biochemical properties of a cytoplasmic cyclomaltodextrinase (neopullulanase) from *Alicyclobacillus acidocaldarius*, reclassification of a group of enzymes, *FEMS Microbiol. Lett.* 183 (2000) 55–61.
- [14] B. Di Lauro, M. Rossi, M. Moracci, Characterization of a beta-glycosidase from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius*, *Extremophiles* 10 (2006) 301–310.
- [15] R. Gul-Guven, K. Guven, A. Poli, B. Nicolaus, Purification and some properties of a beta-galactosidase from the thermoacidophilic *Alicyclobacillus acidocaldarius* subsp. *rittmannii* isolated from Antarctica, *Enzyme Microbiol. Technol.* 40 (2007) 1570–1577.
- [16] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* 72 (1976) 248–254.
- [17] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
- [18] R.J. Leatherbarrow, GraFit version 3.0, Erithacus Software Ltd., Staines, U.K., 1992.
- [19] M. Moracci, A. Trincon, G. Perugino, M. Ciamarella, M. Rossi, Restoration of the activity of active-site mutants of the hyperthermophilic beta-glycosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophiles, *Biochemistry* 37 (1998) 17262–17270.
- [20] T. Yuan, P. Yang, Y. Wang, K. Meng, H. Luo, W. Zhang, N. Wu, Y. Fan, B. Yao, Heterologous expression of a gene encoding a thermostable b-galactosidase from *Alicyclobacillus acidocaldarius*, *Biotechnol. Lett.* (in press), doi:10.1007/s10529-007-9551-y.
- [21] J. Bassias, R. Bruckner, Regulation of lactose utilization genes in *Staphylococcus xylosus*, *J. Bacteriol.* 180 (1998) 2273–2279.
- [22] J. Strey, K.D. Wittchen, F. Meinhardt, Regulation of beta-galactosidase expression in *Bacillus megaterium* DSM319 by a XylS/AraC-type transcriptional activator, *J. Bacteriol.* 181 (1999) 3288–3292.
- [23] S. Shipkowski, J.E. Brechley, Bioinformatic, genetic, and biochemical evidence that some glycoside hydrolase family 42 beta-galactosidases are arabinogalactan type I oligomer hydrolases, *Appl. Environ. Microbiol.* 72 (2006) 7730–7738.

- [24] A. Amine, D. Moscone, R.A. Bernardo, E. Marconi, G. Palleschi, A new enzymatic spectrophotometric assay for the determination of lactulose in milk, *Anal. Chim. Acta* 406 (2000) 217–224.
- [25] M. Mayer, M. Genrich, W. Kunnecke, U. Bilitewski, Automated determination of lactulose in milk using an enzyme reactor and flow analysis with integrated dialysis, *Anal. Chim. Acta* 324 (1996) 37–45.
- [26] D.L. Zechel, S.G. Withers, Dissection of nucleophilic and acid-base catalysis in glycosidases, *Curr. Opin. Chem. Biol.* 5 (2001) 643–649.
- [27] G. Perugino, A. Trincon, A. Giordano, J. van der Oost, T. Kaper, M. Rossi, M. Moracci, Activity of hyperthermophilic glycosynthases is significantly enhanced at acidic pH, *Biochemistry* 42 (2003) 8484–8493.
- [28] E.H. Rydberg, C. Li, R. Maurus, C.M. Overall, G.D. Brayer, S.G. Withers, Mechanistic analyses of catalysis in human pancreatic alpha-amylase: detailed kinetic and structural studies of mutants of three conserved carboxylic acids, *Biochemistry* 41 (2002) 4492–44502.
- [29] B. Cobucci-Ponzano, A. Trincon, A. Giordano, M. Rossi, M. Moracci, Identification of an archaeal alpha-L-fucosidase encoded by an interrupted gene. Production of a functional enzyme by mutations mimicking programmed-1 frameshifting, *J. Biol. Chem.* 278 (2003) 14622–14631.
- [30] C.A. Tarling, S. He, G. Sulzenbacher, C. Bignon, Y. Bourne, B. Henrissat, S.G. Withers, Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from *Thermotoga maritima* through trapping of a covalent glycosyl-enzyme intermediate and mutagenesis, *J. Biol. Chem.* 278 (2003) 47394–47399.



Contents lists available at ScienceDirect

Biochimie

journal homepage: [www.elsevier.com/locate/biochi](http://www.elsevier.com/locate/biochi)

## Research paper

# The molecular characterization of a novel GH38 $\alpha$ -mannosidase from the crenarchaeon *Sulfolobus solfataricus* revealed its ability in de-mannosylating glycoproteins

Beatrice Cobucci-Ponzano<sup>a</sup>, Fiorella Conte<sup>a</sup>, Andrea Strazzulli<sup>a</sup>, Clemente Capasso<sup>a</sup>,  
Immacolata Fiume<sup>a</sup>, Gabriella Pocsfalvi<sup>a</sup>, Mosè Rossi<sup>a,b</sup>, Marco Moracci<sup>a,\*</sup>

<sup>a</sup> Institute of Protein Biochemistry – Consiglio Nazionale delle Ricerche, Via P. Castellino 111, 80131 Naples, Italy

<sup>b</sup> Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli “Federico II”, Complesso Universitario di Monte S. Angelo, Via Cinthia 4, 80126 Naples, Italy

## ARTICLE INFO

## Article history:

Received 28 January 2010

Accepted 29 July 2010

Available online xxx

## Keywords:

Glycobiology

Thermophilic enzymes

Thermostability

Glycoprotein

Mannooligosaccharides

## ABSTRACT

$\alpha$ -Mannosidases, important enzymes in the N-glycan processing and degradation in Eukaryotes, are frequently found in the genome of Bacteria and Archaea in which their function is still largely unknown. The  $\alpha$ -mannosidase from the hyperthermophilic Crenarchaeon *Sulfolobus solfataricus* has been identified and purified from cellular extracts and its gene has been cloned and expressed in *Escherichia coli*. The gene, belonging to retaining GH38 mannosidases of the carbohydrate active enzyme classification, is abundantly expressed in this Archaeon. The purified  $\alpha$ -mannosidase activity depends on a single  $\text{Zn}^{2+}$  ion per subunit is inhibited by swainsonine with an  $\text{IC}_{50}$  of 0.2 mM. The molecular characterization of the native and recombinant enzyme, named Ss $\alpha$ -man, showed that it is highly specific for  $\alpha$ -mannosides and  $\alpha(1,2)$ ,  $\alpha(1,3)$ , and  $\alpha(1,6)$ -mannobioses. In addition, the enzyme is able to demannosylate  $\text{Man}_3\text{GlcNAc}_2$  and  $\text{Man}_7\text{GlcNAc}_2$  oligosaccharides commonly found in N-glycosylated proteins. More interestingly, Ss $\alpha$ -man removes mannose residues from the glycosidic moiety of the bovine pancreatic ribonuclease B, suggesting that it could process mannosylated proteins also *in vivo*. This is the first evidence that archaeal glycosidases are involved in the direct modification of glycoproteins.

© 2010 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

$\alpha$ -D-Mannosidases are key enzymes involved in processing mannosylated glycans and complex-type N-linked oligosaccharides containing mannose in eukaryotic cells [1]. These enzymes are grouped in four glycoside hydrolase families (GH) of the Carbohydrate Active enZYme classification (CAZy) (<http://www.cazy.org/>) [2], namely GH38, GH47, GH76, and GH92. For the last two families, which include enzymes hydrolysing  $\alpha$ -(1,6)-mannan and  $\alpha$ -(1,2),  $\alpha$ -(1,3), and  $\alpha$ -(1,6)-mannosyl-oligosaccharides, respectively, a limited number of reports are available. Instead, extensive studies on GH38 and GH47 have given interesting information on their function,

general folding, catalysed reaction mechanisms, and specificity to inhibitors.

GH47, or class I enzymes, which have a  $(\alpha/\alpha)_7$  fold, are typically inhibited by 1-deoxymannojirimycin, perform hydrolysis by *inverting* the configuration of the anomeric carbon in the product if compared to the substrate, and are specific for the  $\alpha$ -(1,2) bonds in  $\text{Man}_3\text{GlcNAc}_2$  (EC.3.2.1.113) [3–7]. This family includes only bacterial and eukaryotic enzymes. They are involved in early secretory pathway in Endoplasmic Reticulum and Golgi where they perform N-glycans maturation and quality control by producing high-mannose  $\text{Man}_5\text{GlcNAc}_2$  (for reviews see [8,9]).

Enzymes belonging to GH38, also named class II  $\alpha$ -mannosidases, have a multidomain structure consisting of an N-terminal  $\alpha/\beta$  domain, a three-helical bundle, and an all- $\beta$  C-terminal domain [10]. They are typically inhibited by swainsonine and follow a *retaining* reaction mechanism [11,12]. The eukaryotic enzymes from this family are from Golgi, cytosol, and lysosome.

Golgi  $\alpha$ -mannosidases are of biomedical interest as potential anti-cancer targets [13,14]. The most extensively studied Golgi  $\alpha$ -mannosidase II is the enzyme from *Drosophila melanogaster*

Abbreviations: Ss $\alpha$ -man, *Sulfolobus solfataricus*  $\alpha$ -mannosidase; ManA, *Picrophilus torridus*  $\alpha$ -mannosidase; GH, glycoside hydrolase families; 3D, three-dimensional; CD, Circular Dichroism; HPAE–PAD, High-Performance Anion-Exchange chromatography with Pulsed Amperometric Detection; 4-NP- $\alpha$ -D-Man, 4-nitrophenyl- $\alpha$ -D-mannopyranoside; DNJ, 1-deoxymannojirimycin; RLM-RACE, 5' RNA Ligase Mediated Rapid Amplification of cDNA Ends.

\* Corresponding author. Tel.: +39 081 6132271; fax: +39 081 6132277.

E-mail address: [m.moracci@ibp.cnr.it](mailto:m.moracci@ibp.cnr.it) (M. Moracci).

0300-9084/\$ – see front matter © 2010 Elsevier Masson SAS. All rights reserved.  
doi:10.1016/j.biochi.2010.07.016

Please cite this article in press as: B. Cobucci-Ponzano, et al., The molecular characterization of a novel GH38  $\alpha$ -mannosidase from the crenarchaeon ..., Biochimie (2010), doi:10.1016/j.biochi.2010.07.016

(dGMII), for which the subsite specificity and the molecular bases of catalysis and inhibition were revealed by several 3D-structure determinations and kinetic analyses [10,15–17]. This enzyme acts on  $\alpha$ -(1,3), and  $\alpha$ -(1,6) mannosidic bonds (EC.3.2.1.24, –114), which are cleaved sequentially in the same catalytic site. dGMII requires for the reaction the presence of the terminal  $\beta$ -(1,2)-GlcNAc and is involved in the maturation/diversification of hybrid N-glycans, converting GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> into GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub> [10,17]. Interestingly, this enzyme shows three sugar binding sites (namely the catalytic, holding, and anchor sites) occupied by Man<sub>5</sub> ( $\alpha$ -(1,6)-Man), Man<sub>4</sub> ( $\alpha$ -(1,3)-Man), and GlcNAc<sub>3</sub> ( $\beta$ -(1,2)-GlcNAc), respectively, with an essential zinc atom involved in both substrate binding and catalysis [17].

GH38 includes other Golgi special  $\alpha$ -D-mannosidases III, which are also specific for  $\alpha$ -(1,3), and  $\alpha$ -(1,6) mannosidic bonds, but are inactive on GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>. They convert the high-mannose N-glycan Man<sub>5</sub>GlcNAc<sub>2</sub> into Man<sub>3</sub>GlcNAc<sub>2</sub> in an alternate route for the production of hybrid GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub> [18].

Lysosomal GH38  $\alpha$ -D-mannosidases, optimally active at acidic pH, hydrolyse all  $\alpha$ -mannosidic linkages on mannose glycans originating from glycoprotein catabolism to yield Man<sub>1</sub>GlcNAc<sub>(1–2)</sub> [19,20]. Deficiency in lysosomal  $\alpha$ -D-mannosidases causes  $\alpha$ -mannosidosis, a severe genetic disease producing progressive mental retardation in approximately 1 of 500,000 live births [21]. The bovine lysosomal  $\alpha$ -mannosidase (bLAM), which has been crystallized and biochemically characterized, showed low sequence identity to dGMII but an overall structural similarity [22]. The crystallographic study identified the molecular basis of some mutations causing mannosidosis and an interesting mechanism of activation at low pH [22]. It is worth mentioning that in mammals the core structure Man<sub>3</sub>GlcNAc<sub>2</sub> is handled differently in lysosomes [19]. For instance, a core-specific  $\alpha$ -(1,6) mannosidase hydrolyses efficiently Man<sub>3</sub>GlcNAc but not Man<sub>3</sub>GlcNAc<sub>2</sub> requiring the action of a chitobiase for the removal of a GlcNAc at the reducing end [23].

Despite the fact that prokaryotic  $\alpha$ -D-mannosidases represent more than 60% of the total entries from families GH38, GH47, GH76, and GH92, (according to CAZy at the time of the submission of this paper), the enzymes from Bacteria and Archaea are relatively less studied and far less information is available if compared to Eukaryotic  $\alpha$ -D-mannosidases. In families GH47 and GH76 no prokaryotic enzymes have been characterized so far. In the former family only 4 entries have been identified in Bacteria, while, in GH76, although the 104 prokaryotic entries sum to about 50% of the total members of this family, only the 3D-structure of a GH76  $\alpha$ -D-mannosidase from *Listeria innocua* has been deposited (pdb entry 3K7X).

In GH92, in which the vast majority of the enzymes are bacterial,  $\alpha$ -mannosidases have been recently characterized only in *Bacteroides thetaiotaomicron*, a colonic bacterium commonly found in human gut [24]. The activity screening of the GH92 enzymes from this source, and the inspection of the 3D-structure of an  $\alpha$ -(1,2)-mannosidase, allowed clearing the stereochemistry and the catalytic reaction mechanism, indicating that they are involved in the degradation of N-glycans that are present on host and dietary glycoproteins [24].

According to CAZy website <http://www.cazy.org/>, about 74% (~350 open reading frames (ORF) out of ~480) of enzymes belonging to GH38 are from Archaea and Bacteria. However, very few of these  $\alpha$ -mannosidases have been characterized in detail. In *Escherichia coli*, the enzyme encoded by *mngB* gene, is involved in the utilization of the osmolyte 2-O- $\alpha$ -mannosyl-D-glycerate [25]. In *Mycobacterium tuberculosis*,  $\alpha$ -D-mannosidases modify mycobacterial glycoconjugates such as lipoarabinomannans, key cell surface molecules in host–pathogen interactions [26,27]. In *Bacillus* sp., an  $\alpha$ -D-mannosidase is involved in the depolymerisation of the

mannose-containing polymer xanthan that is used as carbon source [28] while the function *in vivo* of an  $\alpha$ -D-mannosidase from the thermophilic bacterium *Thermotoga maritima* (TmGH38) is unknown [29]. More recently, the first 3D-structure of a GH38 prokaryotic  $\alpha$ -mannosidase from the human pathogen *Streptococcus pyogenes* (SpGH38) was reported [30]. SpGH38, revealing similarity with dGMII, is specific for the hydrolysis of  $\alpha$ -(1,3)-mannosidic linkages; the presence of the encoding gene into an operon including other glycoside hydrolases suggested its involvement in N- and, possibly, O-glycan degradation.

Archaeal  $\alpha$ -mannosidases have been identified so far only in families GH38 and GH76. The only report available describes the molecular characterization of the enzyme from GH38 of the extreme acidophilic Euryarchaeon *Picrophilus torridus* (ManA). The enzyme might be involved in the utilization of exogenous glycans and in the turnover of its own glycoconjugates [31]. So far, there have been no reports on the purification of native  $\alpha$ -D-mannosidases from Archaea.

We describe here the identification of the gene encoding a novel  $\alpha$ -D-mannosidase from the hyperthermophilic Crenarchaeon *Sulfolobus solfataricus*, the purification of the enzyme, and the cloning of the gene. The characterization of the recombinant and the native enzymes demonstrates that the phylogenetically distantly related  $\alpha$ -D-mannosidases from *P. torridus* and *S. solfataricus* have different oligomeric structure, substrate specificity, and metal cofactor. The ability of  $\alpha$ -D-mannosidase from *S. solfataricus* in hydrolysing mannosylated N-glycans and the glycosidic moiety of Rnase B suggests its involvement in the maturation of glycosylated proteins in Archaea.

## 2. Materials and methods

### 2.1. Materials

All commercially available substrates were purchased from Sigma–Aldrich, Dextra, and Carbosynth. The GeneTailor Site-Directed Mutagenesis System was from Invitrogen; the synthetic oligonucleotides were from PRIMM (Italy) and Qiagen.

### 2.2. Purification of native $\alpha$ -mannosidase from *S. solfataricus* (nSsa-man)

*S. solfataricus* cells, strain P2, were grown at 80 °C, pH 3.0 as previously described [32] in a minimal salts medium supplemented with yeast extract (0.1%), sucrose (0.1%), and casaminoacids (0.1%). Growth was monitored spectrophotometrically at 600 nm. When the culture reached an A<sub>600</sub> of 0.6 optical densities, cells were harvested by centrifugation at 5000 × g. The resulting cell pellet was thawed, resuspended in 3 mL g<sup>−1</sup> cells of 20 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, 1% (v/v) Triton X-100. Cells were lysated by three cycles of freeze thawing (5 min at −70 °C; 5 min 37 °C), and centrifuged for 30 min at 10,000 × g. The crude extract was then applied at a flow rate of 2.5 mL min<sup>−1</sup> on a High Load 16/10 Q-Sepharose High Performance column (Amersham Biotech) equilibrated in 20 mM phosphate buffer, pH 7.3. At these conditions, the protein did not bind to the column. Active fractions were pooled, equilibrated in 1 M ammonium sulphate, and applied to a HiLoad 26/10 Phenyl Sepharose High performance (Amersham Biotech), which was equilibrated with 20 mM sodium phosphate buffer, pH 7.3; 1 M ammonium sulphate. After washing with 1-column volumes with the loading buffer, the protein was eluted with a two-step gradient of water (0–80%, 2 column volumes; 80–100%, 3 volumes; 100% 2 volumes) at a flow rate of 3 mL min<sup>−1</sup>; the protein eluted in 90% water. Active fractions were pooled, dialyzed against 20 mM sodium phosphate buffer, pH 7.3, and

concentrated by ultrafiltration on an Amicon YM30 membrane (cut off 30,000 Da). After concentration, the sample was loaded on High Load 16/10 Q-Sepharose High Performance column (Amersham Biotech) equilibrated in 20 mM phosphate buffer, pH 7.3. After washing with 2 column volumes with the loading buffer, the protein was eluted with a two-step gradient using 20 mM phosphate buffer, pH 7.3; 1 M NaCl (0–25%, 4 column volumes; 25–100% 1 volume; 100% 2 volumes), at a flow rate of 3 mL min<sup>-1</sup>. The protein eluted in 10% 20 mM phosphate buffer, pH 7.3; 1 M NaCl. Active fractions were pooled, concentrated, and loaded on a Superdex 200 HR 10/30 gel filtration column for FPLC, equilibrated in 20 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, performed as previously described [32]. Active fractions were pooled and concentrated; protein concentration was determined with the method of Bradford [33]. The native enzyme, named nSs $\alpha$ -man, was 95% pure by SDS-PAGE and was stored at 4 °C.

### 2.3. Cloning and expression in *E. coli*

To express the product of the ORF SSO3006 as a fusion of the glutathione S-transferase of *Schistosoma japonicum* in *E. coli*, the gene was amplified by PCR from the genome of *S. solfataricus*, strain P2, by following the procedure previously described [32] and using the following synthetic primers:

5'-3006: 5'-ATCGCGGATCCATGAGAAACATAAAC-3'  
3'-3006: 5'-ATAGTGCATCCCTAACCCCTCACAC-3'

in which the *Bam*HI site is underlined. The amplified product was ligated to the expression vector pGEX-2TK (GE Healthcare) and the insert of the resulting recombinant plasmid, named pGEX-man, was completely resequenced.

The recombinant  $\alpha$ -mannosidase was expressed from pGEX-man/RB791 *E. coli* cells without induction with IPTG. The recombinant enzyme was purified exploiting the GST-tag and the thrombin cleavage onto the matrix as described by the manufacturer (GE Healthcare). Then, contaminating proteins were eliminated by two subsequent incubations for 20 min at 70 °C and 80 °C, each followed by a centrifugation at 10,000  $\times$  g for 20 min at 4 °C and a Superdex 200 HR 10/30 gel filtration column for FPLC, performed as described above. After this procedure the recombinant enzyme, named rSs $\alpha$ -man, was more than 95% pure by SDS-PAGE.

### 2.4. Site-directed mutagenesis

The mutant Asp338Gly was prepared by site-directed mutagenesis from the pGEX-man plasmid, by using the GeneTailor Site-Directed Mutagenesis System and by following the instructions of the manufacturer. The mutagenic oligonucleotides were the following (mismatch is underlined):

D338Gmut,  
5'-CTAATATTTATGGTTACCCGGCACATTGGGTT-3';  
D338Grev,  
5'-CGGGTAACCATAAAATATTAGCAAGCTTCCC-3'

The gene containing the desired mutation was identified by direct sequencing and completely resequenced.

### 2.5. Enzymatic assays

The standard assay for the  $\alpha$ -mannosidase activity was performed in 50 mM sodium phosphate buffer, pH 6.5 at 65 °C. The indicated aryl-glycoside substrates were used at concentrations ranging between 1 and 5 mM in the final volume of 0.2 mL.

Typically, in each assay we used about 0.9–3  $\mu$ g (8–27 pmoles) of enzyme. At appropriate times (typically from 1 to 10 min for 4-NP- $\alpha$ -D-Man, the substrate of the enzyme, and up to 3 h for the other aryl-glycosides), reaction mixtures were transferred in 0.8 mL of an iced solution of 1 M Na<sub>2</sub>CO<sub>3</sub> and then the absorbance was determined at 420 nm ( $\epsilon_{\text{mM}} = 17.2 \text{ mM}^{-1} \text{ cm}^{-1}$ ). One enzymatic unit is defined as the amount of enzyme catalysing the conversion of 1  $\mu$ mole of substrate into product in 1 min, at the indicated conditions.

The activity of the enzymes on  $\alpha$ -(1,2),  $\alpha$ -(1,3), and  $\alpha$ -(1,6)-D-mannobiose substrates (5 mM) was determined at standard conditions for 3 h by measuring the released D-mannose with the K-MANGL kit (Megazyme, Ireland) exploiting the enzymes hexokinase, phosphomannose isomerase, phosphoglucose isomerase, and glucose-6-phosphate dehydrogenase. The amount of NADPH formed in the reaction catalysed by the last enzyme is stoichiometric with the amount of D-mannose.

The activity of nSs $\alpha$ -man on Man<sub>3</sub>GlcNAc<sub>2</sub> and Man<sub>7</sub>GlcNAc<sub>2</sub> oligosaccharides was measured by incubating the native enzyme (10.6  $\mu$ g, 96 pmoles) at standard conditions; aliquots were withdrawn at different times and stored in ice. Reaction products were analysed by High-Performance Anion-Exchange chromatography with Pulsed Amperometric Detection (HPAE–PAD) equipped with a PA200 column (Dionex, USA). Runs were performed in 50 mM NaOH with a linear gradient of 1–8 mM sodium acetate, 0.5 mL min<sup>-1</sup> for 60 min.

### 2.6. Kinetic and inhibition studies

Kinetic constants on 4-NP- $\alpha$ -D-Man were measured at standard conditions by using concentrations of substrate ranging between 0.1 and 6 mM. Typically, in each assay we used about 0.6  $\mu$ g and 1.5  $\mu$ g of nSs $\alpha$ -man and rSs $\alpha$ -man, respectively. For nSs $\alpha$ -man the final volume of the reaction mixture was 1 mL and the increment of absorbance was monitored for 1 min at 405 nm at which 4-nitrophenol showed an extinction coefficient ( $\epsilon_{\text{mM}}$ ) of 9.34 mM<sup>-1</sup> cm<sup>-1</sup> at 65 °C. For rSs $\alpha$ -man, the final volume was 0.2 mL and, after 5 min of incubation (the initial velocity is linear for at least 6 min), reaction mixtures were transferred in 0.8 mL of iced 1 M Na<sub>2</sub>CO<sub>3</sub>. The absorbance was measured at 420 nm ( $\epsilon_{\text{mM}} = 17.2 \text{ mM}^{-1} \text{ cm}^{-1}$ ) at room temperature.

All kinetic data were calculated as the average of at least two experiments and were plotted and refined with the program GraFit [34].

To determine the IC<sub>50</sub> for the inhibitors swainsonine and 1-deoxymannojirimycin (DNJ), 0.8  $\mu$ g nSs $\alpha$ -man (7.3 pmoles) were incubated in the presence of increasing concentrations of the inhibitor (0.05–0.6 mM for swainsonine and 0.1–2 mM for DNJ) in 50 mM sodium phosphate buffer, pH 6.5; 2.5 mM 4-NP- $\alpha$ -D-Man, at 70 °C, in a final volume of 0.2 mL and then assayed with the Na<sub>2</sub>CO<sub>3</sub> method described above. Identical mixtures containing all the reagents with the exception of the enzyme were used as blank.

### 2.7. General characterization

Molecular mass of denatured and native  $\alpha$ -mannosidase enzymes were measured as previously described [32]. Dependence on temperature was determined by assaying 1.5  $\mu$ g (13.5 pmoles) of enzyme in 50 mM sodium phosphate buffer, pH 6.5, in a final volume of 1 mL on 4.5 mM 4-NP- $\alpha$ -D-Man in the temperature range 40–90 °C. The assay was performed in quartz cuvettes with an optical pathlength of 1.0 cm in a Cary 50 BIO UV–Visible thermostated spectrophotometer (Varian, Australia) collecting linear progress curves (times 2–10 min). The  $\epsilon_{\text{mM}}$  extinction coefficients for 4-nitrophenol were accurately measured at each temperature



and were in the range  $6.9\text{--}10.9\text{ mM}^{-1}\text{ cm}^{-1}$ . Thermal stability was tested by incubating pure  $\alpha$ -mannosidase enzymes ( $0.001\text{ mg mL}^{-1}$ ) in  $50\text{ mM}$  sodium phosphate buffer, pH 6.5, at the indicated temperatures. At intervals, aliquots were withdrawn from the mixture, transferred in ice, and assayed at standard conditions at  $65\text{ }^{\circ}\text{C}$  on  $4.5\text{ mM}$  4-NP- $\alpha$ -D-Man substrate in a final volume of  $1\text{ mL}$  as described above.

Circular Dichroism (CD) experiments were carried out using a Jasco-810 spectropolarimeter (Jasco International, Tokyo, Japan). Far-UV CD spectra were recorded over the  $190\text{--}250\text{ nm}$  range at a scanning speed of  $20\text{ nm/min}$  with a  $8\text{ s}$  response, subjected to 5-fold signal averaging. Samples of  $0.1\text{ mg mL}^{-1}$  protein in  $10\text{ mM}$  sodium phosphate buffer (pH 7.0) were measured in a  $0.1\text{ cm}$  quartz cuvette at  $25\text{ }^{\circ}\text{C}$ . Data were processed using Jasco software after removing the residual noise and normalization of the spectra. Buffer scans were recorded under the same conditions and they were subtracted from the protein spectra before analysis.

## 2.8. Mapping the 5' end of *Ssa-man* mRNA

The 5' RNA Ligase Mediated Rapid Amplification of cDNA Ends (RLM-RACE) method [35] was used to determine the transcription initiation site using the 5'/3' RACE Kit, 2nd Generation (Roche), by following the instructions of the manufacturer. Total RNA was isolated from *S. solfataricus* P2 (mid-exponential phase), grown as described above, by the Qiagen RNA extraction system. The gene-specific primers used to transcribe the mRNA into first-strand cDNA and for the first and second PCR amplification round were:

3006SP1 5'-AGTTTGTGACAATTCCTCTCAAA-3' and the nested primer:  
3006SP2 5'-TGAACCATCTACTTACATCCAAT-3'

The PCR products were ligated into the pGEM-T-easy vector (Promega); the resulting plasmids were used to transform *E. coli* DH5 $\alpha$  cells and positive clones were sequenced with the universal T7 and SP6 oligonucleotides.

## 2.9. Real-Time RT-PCR

Total RNA was prepared as described above, extensively digested with DNase (Ambion) and the absence of DNA was assessed by the lack of PCR amplification with each set of primers described below. Total cDNA was obtained using QuantiTect Reverse Transcription Kit (Qiagen). cDNA was then amplified in a BioRad LightCycler using the DyNAmo HS Syber Green qPCR Kit (Finnzymes). Oligonucleotides used for amplification were the following:

Left3006: 5'-ATTGATACCGCTTGGCTTTG-3'  
Right3006: 5'-AGGGCTGCTAATCCCATTT-3'

for the SSO3006 gene and

Left16S: 5'-GAATGGGGGTGATACTGTGCG-3'  
Right16S: 5'-TTTACAGCCGGGACTACAGG-3'

for the 16S rRNA gene. Optimal melting temperatures for each primer pair were determined by performing real-time analysis with a temperature gradient ranging over  $10\text{ }^{\circ}\text{C}$ ; negative controls with no template cDNA were always included. PCR conditions were:  $15\text{ min}$  at  $95\text{ }^{\circ}\text{C}$  for initial denaturation, followed by  $40\text{ cycles}$  of  $10\text{ s}$  at  $95\text{ }^{\circ}\text{C}$ ,  $25\text{ s}$  at  $56\text{ }^{\circ}\text{C}$  and  $35\text{ s}$  at  $72\text{ }^{\circ}\text{C}$ , and a final step of  $10\text{ min}$  at  $72\text{ }^{\circ}\text{C}$ . Product purity was controlled by melting point analysis of setpoints with  $0.5\text{ }^{\circ}\text{C}$  temperature increase from  $72\text{ }^{\circ}\text{C}$  to

$95\text{ }^{\circ}\text{C}$ . PCR products were analysed on  $2\%$  agarose gels and visualized by ethidium bromide staining. Data reported are from two independent RNA preparations; each cDNA was then used for two independent amplifications and in each amplification samples were in triplicate.

## 2.10. Quantification of metals

Metal content was determined by a graphite furnace atomic absorption spectrophotometer (Analyst 800, Perkin–Elmer, Norwalk, CT, USA). The measurements were performed using a Zeeman-effect background correction system. Pyrolytic graphite-coated THGA tube (PerkinElmer) with an integrated L'vov-type platform was used in the metal determinations. Metal concentration values were obtained using a three-point calibration curve. Recombinant holoenzyme was prepared by incubating the enzyme preparation for  $2\text{ h}$  at  $65\text{ }^{\circ}\text{C}$  in  $50\text{ mM}$  sodium phosphate buffer pH 6.5 and  $1\text{ mM}$   $\text{ZnCl}_2$ .

## 2.11. De-mannosylation analysis of Rnase B

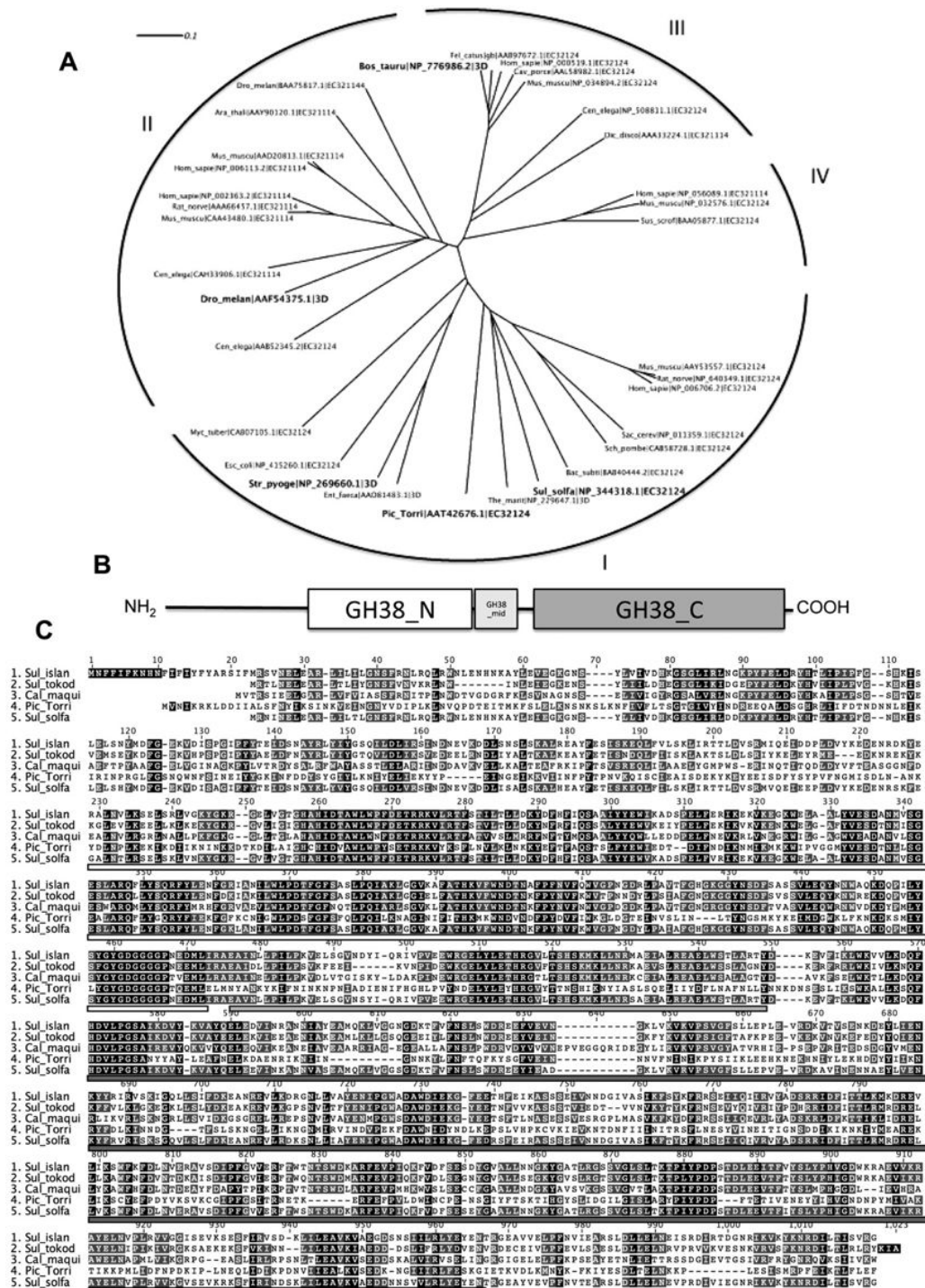
Rnase B (Sigma–Aldrich)  $1.2\text{ nmoles}$  ( $12\text{ }\mu\text{M}$ ) were incubated with  $96\text{ pmoles}$  ( $96\text{ }\mu\text{M}$  corresponding to  $0.03\text{ units}$ ) of nSsa-man in  $10\text{ mM}$  sodium acetate buffer, pH 5.0 and  $2\text{ mM}$   $\text{ZnCl}_2$ , for  $16\text{ h}$  at  $50\text{ }^{\circ}\text{C}$ . Blank sample was prepared in the same way but in the absence of nSsa-man. The reaction mixture was desalted by using ZipTip C4 (Millipore). Proteins were eluted in  $10\text{ }\mu\text{L}$   $50\%$  acetonitrile  $0.1\%$  (v/v) trifluoroacetic acid and spotted ( $1\text{ }\mu\text{L}$ ) on a normal phase ProteinChip array (NP20; Bio-Rad) in triplicates.  $1\text{ }\mu\text{L}$  of sinapinic acid (Bruker Daltonics)  $5\text{ mg mL}^{-1}$  in  $50\%$  acetonitrile  $0.1\%$  (v/v) trifluoroacetic acid was applied on the spots and let them air-dry. A SELDI-TOF (Bio-Rad) mass spectrometer was used to measure the change in the molecular mass distribution of the various known glycoforms of Rnase B upon mannosidase treatment. Data were collected by  $400\text{ Hz}$  sampling rate in the mass range of  $m/z$   $1000\text{--}40,000$  and signal optimization range of  $m/z$   $15,000$ . A total of  $710$  laser shots were taken for an averaged mass spectrum. Three independent measurements were performed on each spots. Instrument was calibrated externally with calmix-3 protein standard (Applied Biosystems).

The samples prepared in the same conditions described above were also analysed by HPAE–PAD equipped with a PA200 column. Runs were performed in  $100\text{ mM}$  NaOH,  $0.5\text{ mL min}^{-1}$  for  $20\text{ min}$ .

## 3. Results

### 3.1. Identification of the $\alpha$ -mannosidase from *S. solfataricus*

In the genome of the thermoacidophilic Crenarchaeon *S. solfataricus*, strain P2, the ORF SSO3006, annotated as a putative  $\alpha$ -mannosidase (EC 3.2.1.24), is  $2907\text{ nt}$  long, encodes for a polypeptide of  $968$  amino acids and, in Carbohydrate Active Enzyme (CAZy) classification, belongs to family GH38 including  $\alpha$ -mannosidases from Archaea, Bacteria and Eukarya [2]. The amino acid sequence of SSO3006 is distantly related to most GH38 enzymes showing very low identity ( $<30\%$ ). Fig. 1A shows the unrooted dendrogram resulting from the alignment of SSO3006 with other  $33$  characterized GH38 enzymes according to CAZy website (<http://www.cazy.org/>). This analysis confirms the previously reported clades: II and III includes the enzymes from Golgi and lysosome, respectively, while clade IV contains lysosomal  $\alpha$ -mannosidases from vertebrates specific for  $\text{Man}_3\text{GlcNAc}$  [23]. SSO3006, together with the other archaeal  $\alpha$ -mannosidase from *P. torridus* (ManA) [31] characterized so far, is located in clade I, which is the most heterogeneous group showing enzymes from several bacteria,



**Fig. 1.** Phylogenetic and structural analysis of SSO3006 – (A) Dendrogram of GH38  $\alpha$ -mannosidases. Characterized members of GH38 were downloaded from CAZY database [2] and multiple sequence alignments were performed by using the program ClustalX [36]; a Phylip tree was generated within this program. The tree was edited and displayed as an unrooted dendrogram by using the program Dendroscope 2.6.1 [37]. The leaves of the tree indicate the organism genus and species information (e.g. *Hom. sapiens* corresponds to *Homo sapiens*), a database accession number, and the EC activities or the notation “3D” if a 3D-structure is deposited. The three enzymes for which the 3D-structure was published, the enzyme ManA from *P. torridus*, and the  $\alpha$ -mannosidase that was expressed and characterized in this paper are in highlighted text (boldface). The curves labelled I–IV highlight the four clades reported for GH38 [23]. (B) Coding region of selected archaeal GH38. The coding region is schematically represented showing the boxes of the amino-terminal, mid and carboxy-terminal Pfam motifs [38]. (C) Multiple sequence alignment of archaeal  $\alpha$ -mannosidases. The sequences aligned with the program ClustalW [36] are shown. The organism genus and species information is the same of above. Highly conserved and similar amino acids are highlighted in black and grey, respectively. The regions corresponding to GH38\_N, GH38\_mid, and GH38\_C are underlined in the alignment by white, light grey, and dark grey bars, respectively.

yeasts, human, rat, and mouse. SSO3006 and ManA are clustered with bacterial orthologs and are probably the result of horizontal gene transfer. The divergence between the two archaeal enzymes occurred very early in the tree, before the divergence between the enzymes from yeasts and vertebrates within Eukaryotes. This explains the very low sequence conservation between SSO3006 and ManA (25% identity).

The two enzymes show similar PFAM signature peptide sequences with two amino- and carboxy-terminal motives of 253 and 389 amino acids, respectively, (termed GH38\_N and GH38\_C), separated by a shorter motif of 73 amino acids (GH38\_mid) (Fig. 1B). This 3-domain organization was found in the 3D-structure of the  $\alpha$ -mannosidase from *Enterococcus faecalis* (pdb entry 3LVT) and in SpGH38 [30]. The distribution of these signatures along a multiple alignment of SSO3006 and ManA with other putative archaeal  $\alpha$ -mannosidases is shown in Fig. 1B. SSO3006 is highly similar to the enzymes from other sulfolobales (sequence identities of 68% and 88% with *Sulfolobus tokodaii* and *Sulfolobus islandicus*, respectively) while it shares 53% identity with *Calditerrivirga maquilensis*. These similarities within the Archaea domain reflect well the profoundly divergent phylogenetic distances between the Crenarchaea and Euryarchaea phyla [39]: sulfolobales and *C. maquilensis* belong to the former, more ancestral, phylum while *P. torridus*, from which ManA originates, is an Euryarchaeon.

3D-Structures are available from all clades but IV with SpGH38, dGMII, and bLAM belonging to clades I, II, and III, respectively, (Fig. 1A) [10,22,30]. However, SSO3006 shows extremely low sequence identity with these enzymes (10%, 5%, and 8%, respectively), precluding reliable structure comparisons. In addition, enzymes in clade I show low similarity and wide functional heterogeneity, ranging from dolichol-oligosaccharide turnover, mannosylated glycan catabolism, and N-glycan degradation [25,26,28,41,42]. This does not allow to easily predicting the specificity and the function of the  $\alpha$ -mannosidase from SSO3006. Therefore, we decided to embark in the molecular characterization of this enzyme.

### 3.2. Analysis of the expression in vivo of the $\alpha$ -mannosidase from *S. solfataricus*

SSO3006 lies upstream to a putative glucose dehydrogenase and a 3-ketoacyl reductase (ORFs SSO3003 and SSO3004, respectively) and it has opposed transcriptional direction with respect to a putative endo-mannanase (SSO3007) and a predicted dehydrogenase (SSO3008) (Fig. 2A). Real-time PCR experiments showed that the *sso3006* gene is transcribed in vivo at about 2000-fold less abundantly than the rRNA 16S. 5' RLM-RACE experiments showed that SSO3006 is expressed by a single transcript starting at an

adenine at position 1 of the first predicted ATG of the gene, indicating that it produces a *leaderless* transcript, lacking a 5' untranslated region (Fig. 2B). This is not uncommon in Archaea (for a review see [43]) and it has been observed also for ManA [31]. Sequence analysis of the upstream gene regions revealed the presence of a canonical TATA-box sequence [40] at position –24 relative to the transcriptional start codon (Fig. 2B).

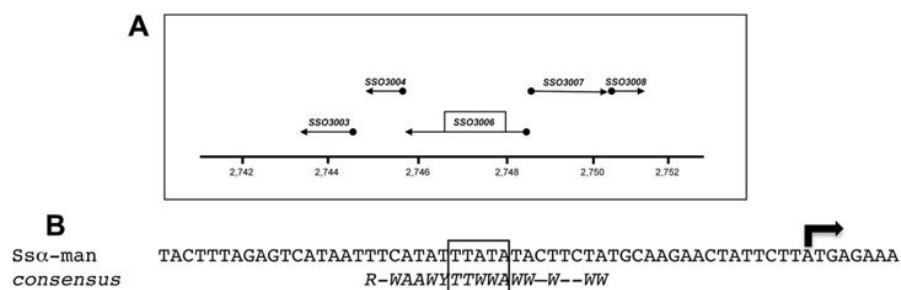
The protein encoded by SSO3006 lacks detectable signal peptide sequences (not shown), suggesting its expression in the cytosol; indeed,  $\alpha$ -mannosidase activity, measured on 4-NP- $\alpha$ -D-Man in fractionated *S. solfataricus* cell extracts was associated with the soluble fraction. The  $\alpha$ -mannosidase was obtained after a 4-steps purification procedure reported in Table 1, which includes two anionic-exchanges, one hydrophobic interaction chromatography, and one gel filtration. In particular, the last step was crucial by improving more than 11-fold the purification of the sample. N-terminal micro-sequencing of the purified protein confirmed that the  $\alpha$ -mannosidase activity found in *S. solfataricus* cell extract is, indeed, the product of the ORF SSO3006 (not shown) and it was named nSs $\alpha$ -man (Fig. 3A). The enzyme represented 0.12% of total proteins *in vivo*.

### 3.3. Characterization of native Ss $\alpha$ -man

The apparent sizes of the enzyme purified from *S. solfataricus* extracts were 110 kDa and 363 kDa in denaturated and native conditions, estimated by SDS-PAGE (Fig. 3A) and by gel filtration chromatography, respectively (Fig. 4A). The calculated molecular weight of the native enzyme suggests a relaxed trimer rather than a tight tetramer.

Atomic absorption measurements identified 2.9 Zn<sup>2+</sup> atoms per trimeric nSs $\alpha$ -man, nicely corresponding to a single atom per subunit. In these conditions, no Cd<sup>2+</sup> was detected, suggesting that nSs $\alpha$ -man bound Zn<sup>2+</sup> *in vivo*. This result demonstrates that the  $\alpha$ -mannosidases from *S. solfataricus* and *P. torridus* have different oligomeric structure and metal dependence: ManA is a dimer, requires Cd<sup>2+</sup> for activity and, in its apo-form, is not activated by Zn<sup>2+</sup> [31]. The presence of metals has not been investigated in ManA, but, presumably, Cd<sup>2+</sup> and not Zn<sup>2+</sup> is structurally bound to this enzyme.

nSs $\alpha$ -man is extremely specific for 4-NP- $\alpha$ -D-Man as no hydrolysis was observed on 4-NP- $\alpha$ -D-glucoside, -galactoside, -xyloside, 4-NP- $\alpha$ -L-arabinoside, -rhamnoside, -fucoside, and 4-NP- $\beta$ -L-fucoside (all assayed at 1 mM final concentration in standard conditions). The steady-state kinetic constants for 4-NP- $\alpha$ -D-Man are shown in Table 2. nSs $\alpha$ -man activity is insensitive to 1-deoxymannojirimycin at concentration up to 2 mM, while swainsonine inhibits the enzyme with an IC<sub>50</sub> of 0.2 mM. The enzyme is active in



**Fig. 2.** Genomic environment of SSO3006 and analysis of its transcriptional expression in vivo – (A) Genomic organization of the *sso3006* locus; (B) genomic sequence flanking the *sso3006* gene: the transcriptional starting site identified by a 5'-RLM-RACE experiment is indicated by an arrow while the putative TATA-box sequence is boxed. The consensus of the promoter region according to [40] is shown, with R = A/G, Y = T/C, and W = A/T.

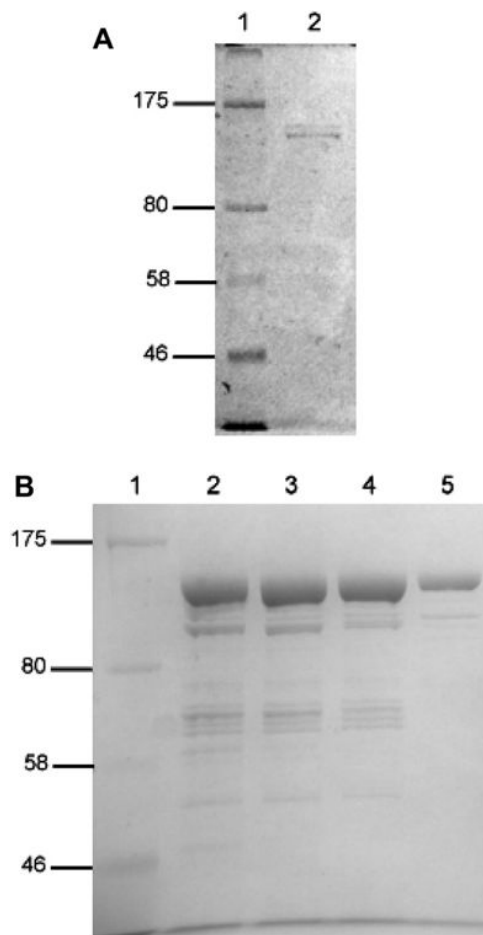


**Table 1**  
Purification of native  $\alpha$ -mannosidase from *S. solfataricus*.

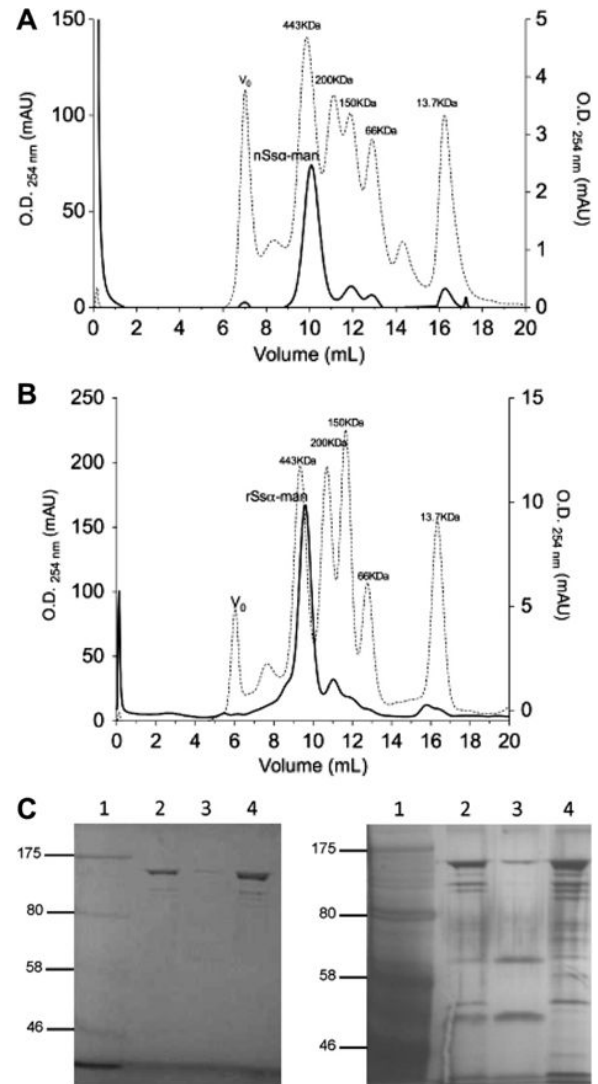
Purification step	Total proteins (mg)	Total units <sup>a</sup> (U)	Specific activity (U/mg)	Purification (fold)	Yields (%)
Cell extract	152	3.04	0.02	1.0	100
Anion-exchange I	65	2.86	0.04	2.0	94
Hydrophobic interaction	2.5	1.50	0.6	30	49
Anion-exchange II	0.16	0.19	1.22	61	6.25
Gel filtration	0.016	0.23	14.4	720	7.56

<sup>a</sup> Assays were performed at standard conditions on 1 mM 4-NP- $\alpha$ -D-Man.

a rather wide range of pHs and its maximal activity (14.4 U/mg) is observed in both 50 mM sodium phosphate, pH 6.5, and in 50 mM sodium acetate, pH 5.0. In addition, as expected for an enzyme from an extreme thermophilic microorganism, nSs $\alpha$ -man was optimally active at temperatures increasing 85 °C (Fig. 5A). Instead, the thermal stability, though higher than that of mesophilic enzymes, was lower than that of other glycoside hydrolases from the same source showing an half-life of 1.5 min at 80 °C (Fig. 5B) [44].



**Fig. 3.** SDS-PAGE analysis of *Ssα-man* – (A) Native enzyme: lanes: 1, Molecular weight markers (kDa); 2, nSs $\alpha$ -man (0.3  $\mu$ g); (B) Recombinant enzyme: lanes: 1, Molecular weight markers (kDa); 2, after affinity chromatography on Glutathione Sepharose 4B (10  $\mu$ g); 3 and 4, after heat treatment at 70 °C (8  $\mu$ g) and 80 °C (7  $\mu$ g), respectively; 5, after gel filtration (2  $\mu$ g).



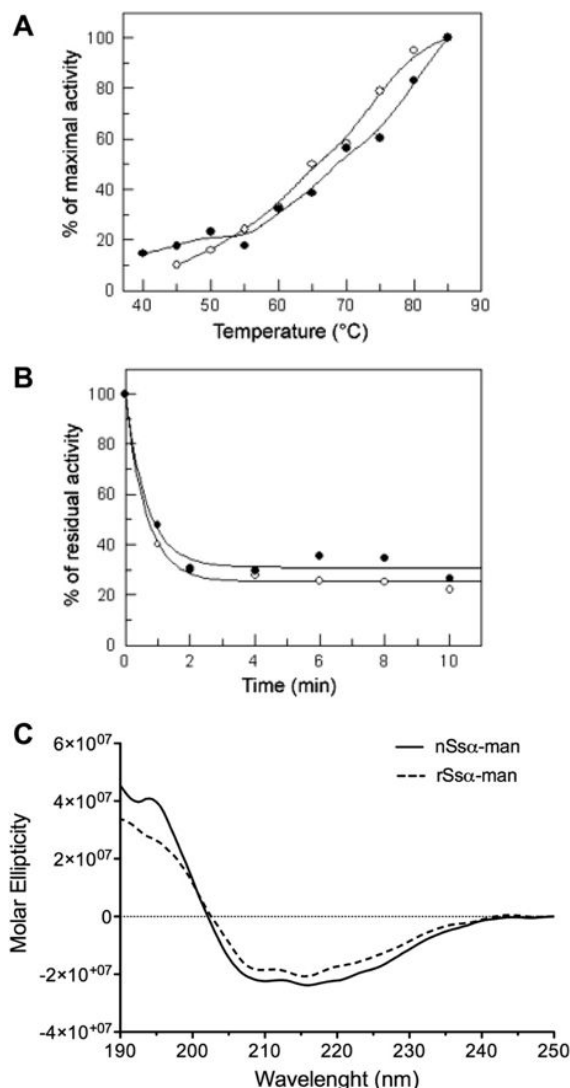
**Fig. 4.** Determination of the native molecular weight of *Ssα-man* – (A) Native enzyme. Elution profile at 280 nm (continuous line) from a Superdex 200 HR 10/30 column in 20 mM phosphate and 150 mM NaCl buffer (pH 7.3). The molecular weight markers (dot lines) are: apoferritin (443,000 Da);  $\beta$ -amylase (200,000 Da); alcohol dehydrogenase (150,000 Da); bovine serum albumin (66,000 Da); ribonuclease A (13,700 Da). (B) Recombinant enzyme. Same conditions as above. (C) SDS-PAGE analysis of fractions of rSs $\alpha$ -man eluted from Superdex 200. Left panel: Coomassie Blue staining; right panel: silver staining. Lanes: 1, Molecular weight markers (kDa); 2, 30  $\mu$ L of fraction eluted at 9.6 mL; 3, 30  $\mu$ L of fraction eluted at 11 mL; 4, 1.5  $\mu$ g of rSs $\alpha$ -man.

### 3.4. Cloning and expression of SSO3006

The SSO3006 gene was cloned by standard molecular biology techniques and was expressed in the *E. coli* strain RB791 as an in-frame fusion at its N-terminal with the Glutathione S-Transferase

**Table 2**  
Steady-state kinetic constant of native and recombinant Ss $\alpha$ -man.

	$k_{cat}$ ( $s^{-1}$ )	$K_M$ (mM)	$k_{cat}/K_M$ ( $mM^{-1} s^{-1}$ )
nSs $\alpha$ -man	$158 \pm 3$	$0.45 \pm 0.04$	351
rSs $\alpha$ -man	$21 \pm 1$	$0.40 \pm 0.09$	54



**Fig. 5.** Activity and stability of *Ssa-man*. (A) Thermal activity of native (empty circles) and recombinant (filled circles) *Ssa-man*. (B) Thermal stability of native (empty circles) and recombinant (filled circles) *Ssa-man* at 80 °C. (C) Far-UV CD spectra of native and recombinant *Ssa-man*.

(GST) purification tag; other trials using different *E. coli* strains gave no or very low expression of the gene. Interestingly, in all cases tested exploiting IPTG-inducible promoters, in the presence of IPTG no expression could be observed and r*Ssa-man* could be expressed only in the absence of inducer. Host cells grew normally excluding toxic effects of the recombinant protein; therefore, the reason of this phenomenon is still obscure. However, it is worth mentioning that Angelov and collaborators report an identical observation for ManA [31]. No other similar reports have been made on  $\alpha$ -mannosidases from other sources, suggesting that this is a peculiarity of archaeal enzymes.

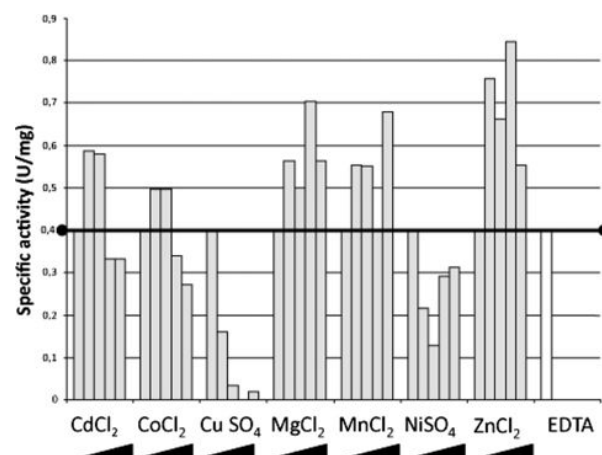
### 3.5. Purification and characterization of r*Ssa-man*

Recombinant *Ssa-man* was obtained in pure form (>95% estimated by SDS-PAGE) by exploiting the GST-tag affinity purification

followed by two heat-precipitations at 70 °C and 80 °C, and a gel filtration chromatography (Fig. 3B). The molecular weights of r*Ssa-man* in native and denatured conditions were identical to those of n*Ssa-man* (Fig. 4). In ManA, a fraction of approximately 120 kDa, observed in gel filtration experiments, was explained as single subunits [31]. In r*Ssa-man*, this fraction is inactive and results from protein contaminants observed by SDS-PAGE with Coomassie and silver staining (Fig. 4C).

The specific activity of r*Ssa-man* measured at the same conditions of the native enzyme on 4-NP- $\alpha$ -D-Man was 36-fold lower (0.4 vs 14.4 U/mg). The activity of r*Ssa-man* was measured in the presence of four concentrations in the 1  $\mu$ M–1 mM range of different metals to test if they can reactivate the enzyme. The highest activation (2-fold) was observed at 0.1 mM ZnCl<sub>2</sub>, followed by MgCl<sub>2</sub> (0.1 mM), MnCl<sub>2</sub> (1 mM), CdCl<sub>2</sub>, and CoCl<sub>2</sub> (both at 10  $\mu$ M and 0.1 mM concentrations). No changes in the enzymatic activity were observed with Cu<sup>2+</sup>, Ni<sup>2+</sup>, and EDTA (Fig. 6). This experiment clearly shows that the activity of *Ssa-man* depends on metal cofactors. Even when the enzyme is assayed in the presence of 0.1 mM ZnCl<sub>2</sub> the specific activity of r*Ssa-man* is still 18-fold lower than that of the n*Ssa-man*. Kawar and co-workers reported that the specific activity of an  $\alpha$ -mannosidase from insect cells increased 10-fold after 2 h incubation with Co<sup>2+</sup> at the temperature of the assay [18]. The incubation of r*Ssa-man* for 2 h in the standard assay mixture in the presence of 1 mM ZnCl<sub>2</sub> allowed to increase the specific activity of the enzyme to 1.6 U/mg on 4.5 mM 4-NP- $\alpha$ -D-Man, therefore, we pre-incubated r*Ssa-man* in this way for all the subsequent characterization.

The stability at 80 °C, the thermal activity, and the overall secondary structure analysed by far-UV CD spectra of r*Ssa-man* are very similar to those of the native enzyme (Fig. 5). Instead, although the affinity for the substrate of the two enzymes was identical within the experimental error, the  $k_{cat}$  and  $k_{cat}/K_M$  on 4-NP- $\alpha$ -D-Man for the recombinant enzyme were lower than those of n*Ssa-man* (Table 2). Atomic absorption measurements on purified r*Ssa-man* identified 2.3 atoms of Zn<sup>2+</sup> per trimer, suggesting that the recombinant enzyme binds slightly less cofactor than n*Ssa-man* (2.9 atoms per trimer). Presumably, this occurs during its heterologous expression; attempts to increase the catalytic efficiency of r*Ssa-man* by adding ZnCl<sub>2</sub> to the growth medium were unsuccessful.



**Fig. 6.** Effect of metal cofactors on r*Ssa-man* activity – The  $\alpha$ -mannosidase activity was measured at standard conditions on 4.5 mM 4-NP- $\alpha$ -D-Man in the presence of five concentrations of metal cofactors (0, 1, 10, 100  $\mu$ M, and 1 mM) and in 0 and 1 mM EDTA.

### 3.6. Identification of the catalytic amino acids

GH38  $\alpha$ -mannosidases cleave the substrate bond with net retention of the anomeric configuration and use a two-step mechanism involving the formation of a covalent mannosyl-intermediate. Two carboxylic acids in the active site act as nucleophile and as general acid/base catalysts, respectively [46]. The nucleophile has been experimentally identified in the  $\alpha$ -mannosidases from jack bean, bovine kidney, and fruit fly [12,15,47]. The multiple amino acid sequence alignment of GH38  $\alpha$ -mannosidases allowed us to predict that in *Ss* $\alpha$ -man Asp338 and Asp429 are the nucleophile and the acid/base of the reaction, respectively (Fig. 7). The acid/base Asp341 in *D. melanogaster* was identified experimentally by site-directed mutagenesis; however, the assignment of this residue in GH38 enzymes is uncertain as it falls in a region with low sequence identity (Fig. 7B) [15]. The *Ss* $\alpha$ -man Asp338 is invariant and corresponds to the catalytic nucleophile experimentally determined. To proof the function of Asp338, we prepared the mutant *Ss* $\alpha$ -manA-sp338Gly that was expressed and purified from *E. coli*. The mutant showed no activity as expected for an enzyme in which the essential catalytic residue was deleted. In addition, when assayed at standard conditions and in the presence of 1 M sodium formate, the activity of the enzyme was partially rescued by this external nucleophile ( $3.5 \times 10^{-4} \pm 0.9 \times 10^{-4}$  U/mg vs  $2.5 \pm 0.7$  U/mg for the wild type enzyme). The chemical rescue of the enzymatic activity of a mutant is one of the approaches used to assign the function of the catalytic residues [46]. Unfortunately, no trans-mannosylation product could be observed by thin layer chromatography analysis of the reaction mixtures, precluding the possibility of determine the structure of the products; however, our data strongly indicate that Asp338 is the nucleophile of the reaction of *Ss* $\alpha$ -man.

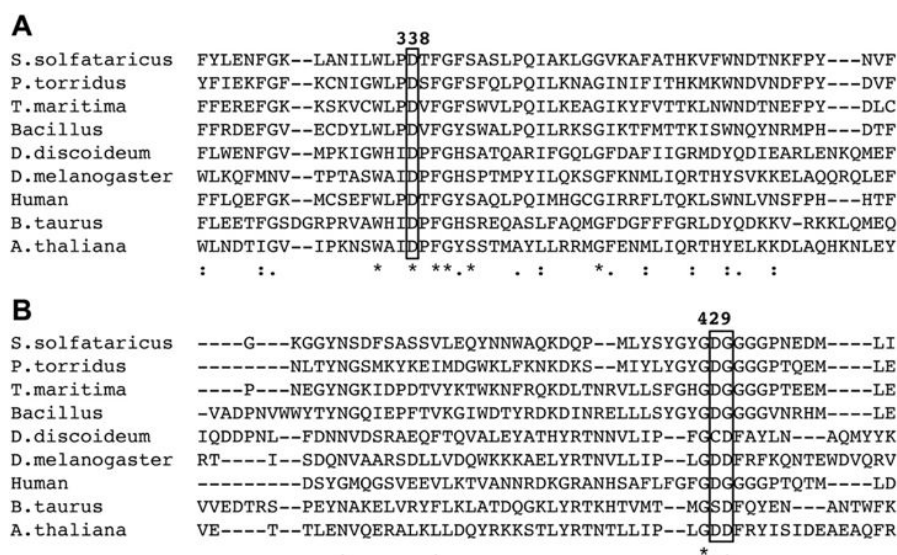
### 3.7. Characterization of the substrate specificity of *Ss* $\alpha$ -man on mannosylated glycans

To understand the possible role *in vivo* of *Ss* $\alpha$ -man we analysed the ability of the enzyme to hydrolyse mannosylated glycans found

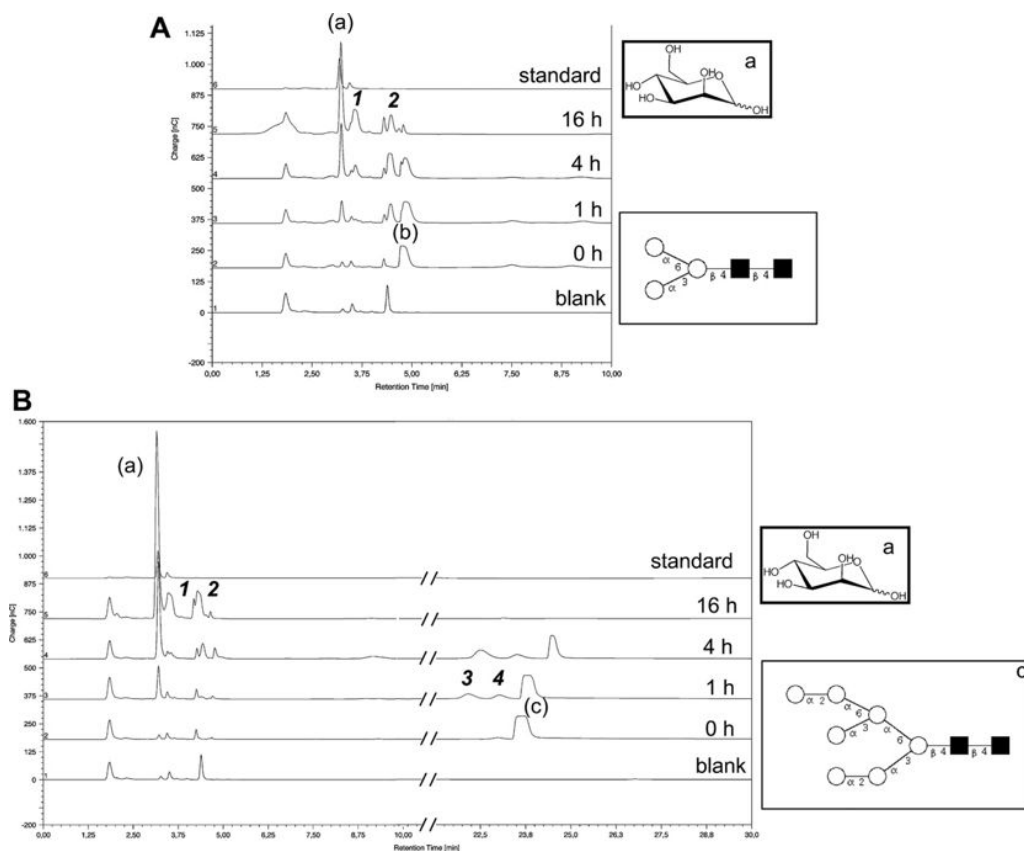
in N-glycosylated proteins. After 3 h of incubation, *rSs* $\alpha$ -man released mannose from  $\alpha(1,3)$ ,  $\alpha(1,6)$ -D-mannopentaose (approximately 0.11 U/mg) and hydrolysed  $\alpha(1,2)$ -,  $\alpha(1,3)$ -, and  $\alpha(1,6)$ -D-mannobiose with specific activities of 0.71 U/mg, 0.45 U/mg, 0.16 U/mg, respectively. Apparently, *Ss* $\alpha$ -man shows lower activity on  $\alpha(1,6)$ -D-mannobiose while the activity of ManA, assayed by others at similar conditions, resulted similar on these substrates [31].

Interestingly, *nSs* $\alpha$ -man was able to recognize as substrates two high-mannose oligosaccharides commonly found in N-glycosylated proteins, namely Man<sub>3</sub>GlcNAc<sub>2</sub> and Man<sub>7</sub>GlcNAc<sub>2</sub> oligosaccharides (Fig. 8). The analysis by HPAE–PAD of the reaction mixtures containing Man<sub>3</sub>GlcNAc<sub>2</sub> showed that, already after 1 h of incubation, free mannose (a) and two well resolved peaks (1, and 2) with retention times lower than that of the substrate (b) could be identified (Fig. 8A). After 16 h of incubation, Man<sub>3</sub>GlcNAc<sub>2</sub> was completely hydrolysed and the three products accumulated in the reaction. In the case of Man<sub>7</sub>GlcNAc<sub>2</sub> (Fig. 8B), after 1 h of incubation, we observed two new peaks (3, and 4) at retention times of ~22–23 min that, together with the substrate (c), progressively disappear with the formation of mannose (a) and of the reaction products 1, and 2. The limited scale of the reaction precluded the structural determination of the observed product, thus, without suitable markers it is difficult to predict the structure of compounds 1–4. Presumably, one of the products (1 or 2) could be Man<sub>1</sub>. GlcNAc<sub>2</sub> containing a  $\beta$ -anomeric bond that cannot be hydrolysed by *Ss* $\alpha$ -man. Instead, the retention times of 3 and 4, which are lower than that of the substrate, might be oligosaccharides resulting from the partial demannosylation of Man<sub>7</sub>GlcNAc<sub>2</sub>. The reduced specificity of *Ss* $\alpha$ -man for  $\alpha(1,6)$ -D-mannobiose might suggest that Man<sub>2</sub>GlcNAc<sub>2</sub> containing a  $\alpha(1,6)$ -D-mannobiose accumulates in the reaction. Further experiments are required to confirm this hypothesis.

The ability of the enzyme of hydrolysing high-mannose oligosaccharides prompted us to test *nSs* $\alpha$ -man on glycoproteins. Ribonuclease B (Rnase B) from bovine pancreas was chosen as a model substrate because it possesses a single N-glycosylation site with high-mannose oligosaccharides Man<sub>5</sub>–Man<sub>9</sub> [48]. Remarkably, the



**Fig. 7.** Multiple sequence alignment of GH38  $\alpha$ -mannosidases – The regions corresponding to the catalytic nucleophile and the acid/base of the reaction (*Ss* $\alpha$ -man numbering), shown in (A) and (B), respectively, were extracted from a sequence alignment performed with the program T-Coffee [45] by using the default settings. The catalytic residues are boxed, the symbols “\*”, “.” and “:” indicate identical, conserved, and semi-conserved residues, respectively. In (B) the acid/base residue is poorly conserved in GH38, and the box encompasses two residues to include the Asp341 in *D. melanogaster* [15].



**Fig. 8.** HPAE–PAD of high-mannose oligosaccharide incubated with Ssa-man – The chromatographic runs of the mixtures containing nSsa-man and Man<sub>3</sub>GlcNAc<sub>2</sub> or Man<sub>7</sub>GlcNAc<sub>2</sub> are shown in (A) and (B), respectively. The incubation times are also indicated on each run. Mannose (a), Man<sub>3</sub>GlcNAc<sub>2</sub> (b), and Man<sub>7</sub>GlcNAc<sub>2</sub> (c) were used as markers, and the blank mixture contains the enzyme only. In the schematic representation of Man<sub>3</sub>GlcNAc<sub>2</sub> and Man<sub>7</sub>GlcNAc<sub>2</sub> mannose and N-acetyl-glucosamine are reported as empty circles and filled squares, respectively, and the anomeric bonds are also indicated.

analysis by SELDI-TOF MS of the reaction mixture containing nSsa-man and Rnase B clearly shows the formation of a new peak at  $m/z$  14,731 next to the peak at  $m/z$  14,892 corresponding to the most abundant Rnase B/Man<sub>5</sub> glycoform (Fig. 9A). The difference in mass between the two peaks corresponds to that of a single mannose residue and fits well with those experimentally identified in the same SELDI-TOF MS spectrum for the different Rnase B glycoforms Man<sub>5</sub>/Man<sub>6</sub>/Man<sub>7</sub>/Man<sub>8</sub>/Man<sub>9</sub>. Therefore, this peak can be attributed to a novel Man<sub>4</sub> resulting from the single deglycosylation of Rnase B/Man<sub>5</sub> or, alternatively, from subsequent deglycosylations of Rnase B/Man<sub>6</sub>–Man<sub>9</sub>. A more detailed characterization of the deglycosylated Rnase B goes beyond the aim of this study and it will be subject of further experiments; however, the analysis by HPAE–PAD of the same reaction mixture allowed us to detect free mannose after 16 h of incubation, confirming that the enzyme is able to release mannose from Rnase B (Fig. 9B). These data indicate that Ssa-man could be putatively involved in the turnover and/or the maturation of glycoprotein also *in vivo*.

#### 4. Discussion

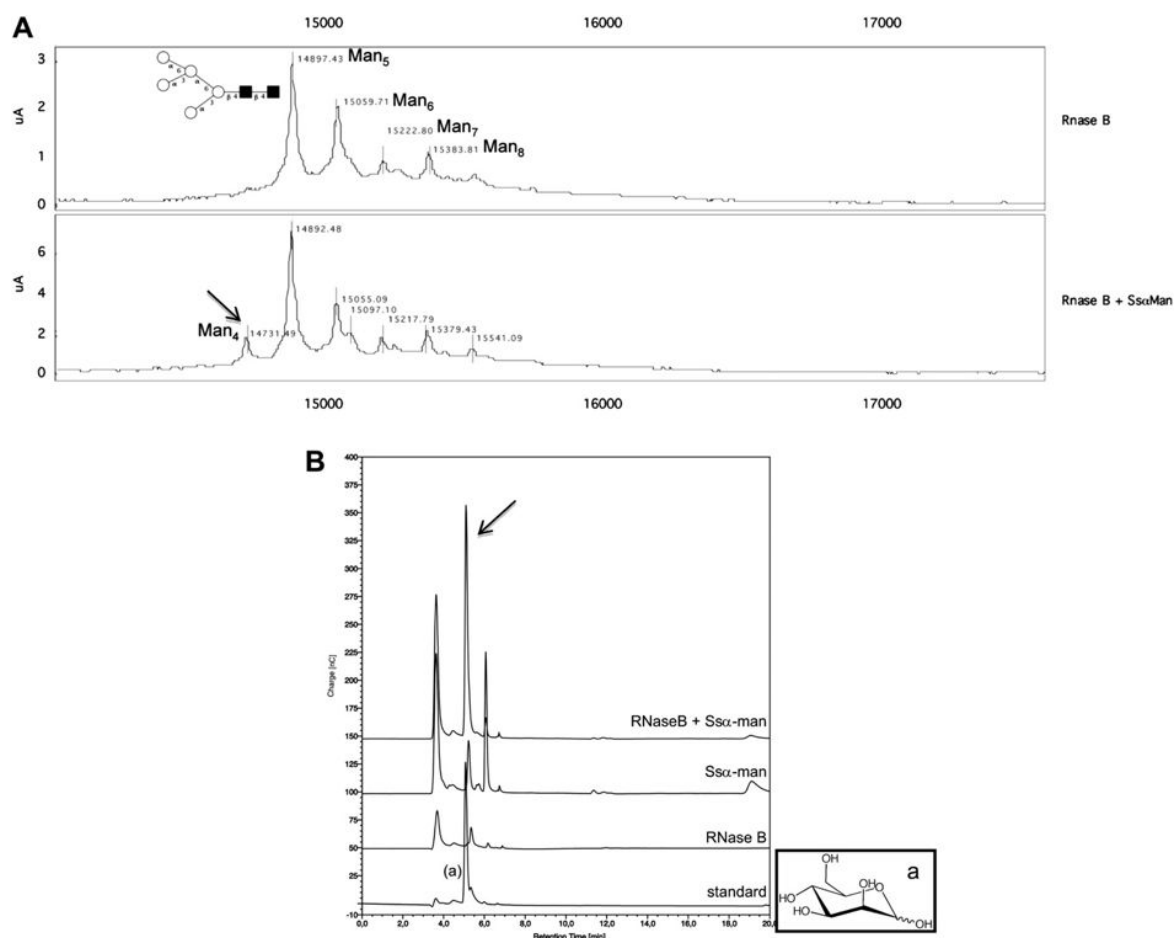
Here we show the molecular characterization of a novel GH38  $\alpha$ -mannosidase from the thermoacidophilic Crenarchaeon *S. solfataricus* (Ssa-man), which is able to hydrolyse high-mannose oligosaccharides and mannosylated proteins. This is the second

archaeal  $\alpha$ -mannosidase characterized so far, and the first from Crenarchaea. The only other known archaeal  $\alpha$ -mannosidase (ManA) has been identified and characterized in the Euryarchaeon *P. torridus* [31]. Our study shows that the characteristics of the two enzymes are rather different. Both  $\alpha$ -mannosidases are intracellular, are expressed as single, leaderless, transcripts *in vivo*, are extremely selective for  $\alpha$ -D-mannosides, and are both inhibited by swainsonine and insensitive to DNJ. Therefore, both can be grouped in class II  $\alpha$ -mannosidases. However, ManA and Ssa-man have only 25% sequence identity, diverging early in the GH38 phylogenetic tree, and, possibly, deriving from a lateral gene transfer.

The two enzymes show diverse oligomeric status and bind different metal cofactors: ManA is a dimer whose activity depends on Cd<sup>2+</sup> while for Ssa-man, which is activated by Zn<sup>2+</sup>, gel filtration experiments suggest that it is a trimer. This is not unusual among  $\alpha$ -mannosidases: the GH38 enzyme from *Bacillus*, 29% identical to Ssa-man, is trimeric ([28] as well as that from *Ginkgo biloba*), whose family is unknown [49]. Several oligomeric status have been found with dimeric (TmGH38, SpGH38, ManA [29–31]), tetrameric (rat, pig, and human [50–52]), and hexameric (yeast [53]) enzymes.

Also metal dependence in GH38 enzymes is rather diverse: cadmium activates the  $\alpha$ -mannosidase from *T. maritima* [29], Co<sup>2+</sup> is the preferred cofactor for the  $\alpha$ -mannosidases from insect and *Bacillus* sp. [18,28], while the activity of the enzymes from jack bean [47], dGMII, LAM, and SpGH38 is strictly dependent on Zn<sup>2+</sup>





**Fig. 9.** *Sα-man demannosylate Rnase B* – (A) SELDI-TOF mass spectrometer analysis of Rnase B incubated with *Sα-man*. Reaction mixtures containing ribonuclease B (Rnase B) from bovine pancreas alone or together with *nSα-man* are reported in the upper and lower part, respectively. In the upper part, the peaks corresponding to  $\text{Man}_5$ – $\text{Man}_8$  are labelled correspondingly and the schematic structure of  $\text{Man}_5$ , represented as described in the legend of Fig. 7, is reported. In the lower part, the new peak at  $m/z$  14,731 is labelled with  $\text{Man}_4$  and highlighted with an arrow. (B) HPAE–PAD analysis of Rnase B incubated with *Sα-man*. The samples analysed by HPAE–PAD were prepared at the same conditions described for those analysed by SELDI-TOF (see Materials and methods). The peak observed in the sample containing both Rnase B and *nSα-man*, showing the same retention time of D-mannose standard (a), is highlighted with an arrow.

[10,22,30]. In this last group of enzymes, this metal is particularly important being located in the  $\alpha/\beta$  portion in which forms an integral part of the  $-1$  subsite of the catalytic site. Here, zinc coordinates the active site nucleophile and the  $\text{Man}_5$  in the  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  substrate [15,17]. The importance of  $\text{Zn}^{2+}$  in catalysis and substrate binding is also confirmed in *Sα-man*. The recombinant enzyme that showed similar secondary structure, oligomeric status, and stability than the enzyme purified from *S. solfataricus*, shows a 7.5-fold lower  $k_{\text{cat}}$  and presents only 2.3 zinc atoms per trimer, suggesting that the impaired activity results from the reduced efficiency in binding the cofactor.

Eukaryotic enzymes from GH38 are mainly involved in the maturation of glycoproteins during the transit in Golgi and ER or in their turnover in the lysosome. The function of cytosolic enzymes is less certain, but they are probably involved in protein recognition and signalling. The function of prokaryotic members has been less studied. The  $\alpha$ -mannosidase from *Bacillus* sp. participates in the degradation of xanthan, an exopolysaccharide of bacterial origin, allowing its exploitation as carbon source [28]. Apart from xanthan, not many exopolysaccharides contain  $\alpha$ -mannose; therefore, the

possibility that the cytosolic *Sα-man* could be involved *in vivo* in glycan degradation as a possible carbon source is questionable. In addition, 2- $\alpha$ -mannosyl-D-glycerate, a compatible solute very common in Euryarchaeota, is absent in *Sulfolobales*, thus, the involvement of *Sα-man* in its metabolism can be ruled out.

The  $\alpha$ -mannosidase from the metanabolism can be ruled out. The  $\alpha$ -mannosidase from the catabolism of the cell wall glycoproteins and glycolipids whose terminal  $\alpha(1,2)$  mannose residues could function as virulence factors [26].  $\alpha$ -Mannose has been identified also in the cell envelope of other human parasites, but it is not a common component of bacterial glycoproteins therefore,  $\alpha$ -mannosidases from bacteria hosted in the human body might be involved in the catabolism of  $\alpha$ -mannosylated glycoconjugates, rather than in their maturation [54,55]. The function of the intracellular enzyme from another pathogen, *S. pyogenes*, is still unknown, however, it might be involved in the degradation of host glycans transported inside the bacterium [30].

The study of protein glycosylation in Archaea is still in its infancy, but it is arousing high interest. In this domain, this process is considered a simpler version of eukaryal N-glycosylation (for

reviews see [56,57]). Briefly, Archaea exploit dolichol-phosphate like Eukaryotes instead of the bacterial undecaprenyl-pyrophosphate lipid carrier. In addition, the single subunit oligosaccharide transferase in Archaea is the simplified version of the homologous multimeric eukaryal counterpart. Nevertheless, many aspects of protein glycosylation in Archaea are obscure: the glycosylation mechanism, the complete set of genes involved and their function, and the structure of the glycosidic component are mostly unknown. However,  $\alpha$ -mannose has been frequently identified in glycosylated proteins and in exopolysaccharides in several Archaea including *S. solfataricus*, *Methanosarcina acetivorans* and *Methanosarcina maizei* [58–60]. Therefore,  $\alpha$ -mannosidases from this source could be involved not only in the catabolism of exogenous glycans, but also in the maturation of nascent glycoproteins. Our study supports this conclusion. *Ss* $\alpha$ -Man purified from *S. solfataricus* extracts shows wide specificity for  $\alpha$ (1,2),  $\alpha$ (1,3), and  $\alpha$ (1,6)-D-mannobiose substrates. In addition, it is able to hydrolyse high-mannose oligosaccharides such as Man<sub>3</sub>GlcNAc<sub>2</sub> and Man<sub>7</sub>GlcNAc<sub>2</sub> and, remarkably, to partially demannosylate Rnase B.

GH38 collects a large variety of  $\alpha$ -mannosidases showing exquisite substrate specificity that has been described in detail giving clear indication on their function. In particular, the availability of the glycans that are their natural substrates *in vivo* allowed probing substrate specificities by 3D-structure and kinetic analyses. The characterization of the substrate specificity of *Ss* $\alpha$ -man is currently hampered by the limited information available on its natural substrates, precluding detailed functional studies. However, the abundant expression of *Ss* $\alpha$ -man *in vivo* and its ability in protein demannosylation makes this enzyme an interesting subject of study to further understand its function *in vivo*.

## Acknowledgements

This work was supported by the project MoMa n. 1/014/06/0 of the Agenzia Spaziale Italiana. We thank Dr. Anna Padula for the 5'RACE experiments and Dr. Federica Santoro for Real-Time PCR experiments.

## References

- [1] A. Herscovics, Importance of glycosidases in mammalian glycoprotein biosynthesis, *Biochim. Biophys. Acta* 1473 (1999) 96–107.
- [2] B.L. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycomics, *Nucleic Acids Res.* 37 (2009) D233–D238.
- [3] K.W. Moremen, R.B. Trimble, A. Herscovics, Glycosidases of the asparagine-linked oligosaccharide processing pathway, *Glycobiology* 4 (1994) 113–125.
- [4] L.O. Tremblay, A. Herscovics, Characterization of a cDNA encoding a novel human Golgi alpha 1,2-mannosidase (IC) involved in N-glycan biosynthesis, *J. Biol. Chem.* 275 (2000) 31655–31660.
- [5] F. Vallee, K. Karavag, A. Herscovics, K.W. Moremen, P.L. Howell, Structural basis for catalysis and inhibition of N-glycan processing class I alpha 1,2-mannosidases, *J. Biol. Chem.* 275 (2000) 41287–41298.
- [6] W. Tempel, K. Karavag, Z.J. Liu, J. Rose, B.C. Wang, K.W. Moremen, Structure of mouse Golgi alpha-mannosidase IA reveals the molecular basis for substrate specificity among class 1 (family 47 glycosylhydrolase) alpha 1,2-mannosidases, *J. Biol. Chem.* 279 (2004) 29774–29786.
- [7] K. Karavag, A. Siriwardena, W. Tempel, Z.J. Liu, J. Glushka, B.C. Wang, K.W. Moremen, Mechanism of class 1 (glycosylhydrolase family 47) alpha-mannosidases involved in N-glycan processing and endoplasmic reticulum quality control, *J. Biol. Chem.* 280 (2005) 16197–16207.
- [8] S.W. Mast, K.W. Moremen, Family 47 alpha-mannosidases in N-glycan processing, *Glycobiology* 415 (2006) 31–46.
- [9] G.Z. Lederkremer, Glycoprotein folding, quality control and ER-associated degradation, *Curr. Opin. Struct. Biol.* 19 (2009) 515–523.
- [10] J.M.H. van den Elsen, D.A. Kuntz, D.R. Rose, Structure of Golgi alpha-mannosidase II: a target for inhibition of growth and metastasis of cancer cells, *EMBO J.* 20 (2001) 3008–3017.
- [11] P.R. Dorling, C.R. Huxtable, S.M. Colegate, Inhibition of lysosomal alpha-mannosidase by swainsonine, an indolizidine alkaloid isolated from *Swainsona canescens*, *Biochem. J.* 191 (1980) 649–651.
- [12] S. Numao, S.M. He, G. Evjen, S. Howard, O.K. Tollersrud, S.G. Withers, Identification of Asp197 as the catalytic nucleophile in the family 38 alpha-mannosidase from bovine kidney lysosomes, *FEBS Lett.* 484 (2000) 175–178.
- [13] M. Granovsky, J. Fata, J. Pawling, W.J. Muller, R. Khokha, J.W. Dennis, Suppression of tumor growth and metastasis in Mgat5-deficient mice, *Nat. Med.* 6 (2000) 306–312.
- [14] H.B. Guo, I. Lee, M. Kamar, M. Pierce, N-Acetylglucosaminyltransferase V expression levels regulate cadherin-associated homotypic cell-cell adhesion and intracellular signaling pathways, *J. Biol. Chem.* 278 (2003) 52412–52424.
- [15] S. Numao, D.A. Kuntz, S.G. Withers, D.R. Rose, Insights into the mechanism of *Drosophila melanogaster* Golgi alpha-mannosidase II through the structural analysis of covalent reaction intermediates, *J. Biol. Chem.* 278 (2003) 48074–48083.
- [16] W. Zhong, D.A. Kuntz, B. Ernber, H. Singh, K.W. Moremen, D.R. Rose, G.J. Boons, Probing the substrate specificity of Golgi alpha-mannosidase II by use of synthetic oligosaccharides and a catalytic nucleophile mutant, *J. Am. Chem. Soc.* 130 (2008) 8975–8983.
- [17] N. Shah, D.A. Kuntz, D.R. Rose, Golgi alpha-mannosidase II cleaves two sugars sequentially in the same catalytic site, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 9570–9575.
- [18] Z. Kavar, K. Karavag, K.W. Moremen, D.L. Jarvis, Insect cells encode a class II alpha-mannosidase with unique properties, *J. Biol. Chem.* 276 (2001) 16335–16340.
- [19] P.F. Daniel, B. Winchester, C.D. Warren, Mammalian alpha-mannosidases—multiple forms but a common purpose, *Glycobiology* 4 (1994) 551–566.
- [20] Y.F. Liao, A. Lal, K.W. Moremen, Cloning, expression, purification, and characterization of the human broad specificity lysosomal acid alpha-mannosidase, *J. Biol. Chem.* 271 (1996) 28348–28358.
- [21] D. Malm, O. Nilssen, Alpha-mannosidosis, *Orphanet J. Rare Dis.* 3 (2008).
- [22] P. Heikinheimo, R. Helland, H.K.S. Leiros, I. Leiros, S. Karlens, G. Evjen, R. Ravelli, G. Schoehn, R. Ruigrok, O.K. Tollersrud, S. McSweeney, E. Hough, The structure of bovine lysosomal alpha-mannosidase suggests a novel mechanism for low-pH activation, *J. Mol. Biol.* 327 (2003) 631–644.
- [23] C. Park, L. Meng, L.H. Stanton, R.E. Collins, S.W. Mast, X.B. Yi, H. Strachan, K.W. Moremen, Characterization of a human core-specific lysosomal alpha 1,6-mannosidase involved in N-glycan catabolism, *J. Biol. Chem.* 280 (2005) 37204–37216.
- [24] Y.P. Zhu, M.D.L. Suits, A.J. Thompson, S. Chavan, Z. Dinev, C. Dumon, N. Smith, K.W. Moremen, Y. Xiang, A. Siriwardena, S.J. Williams, H.J. Gilbert, G.J. Davies, Mechanistic insights into a Ca<sup>2+</sup>-dependent family of alpha-mannosidases in a human gut symbiont, *Nat. Chem. Biol.* 6 (2010) 125–132.
- [25] M.M. Sampaio, F. Chevance, R. Dippel, T. Eppler, A. Schlegel, W. Boos, Y.J. Lu, C.O. Rock, Phosphotransferase-mediated transport of the osmolyte 2-O-alpha-mannosyl- $\alpha$ -glycerate in *Escherichia coli* occurs by the product of the *mngA* (hrsA) gene and is regulated by the *mngR* (farR) gene product acting as repressor, *J. Biol. Chem.* 279 (2004) 5537–5548.
- [26] C.A. Rivera-Marrero, J.D. Ritzenthaler, J. Roman, K.W. Moremen, Molecular cloning and expression of an alpha-mannosidase gene in *Mycobacterium tuberculosis*, *Microb. Pathog.* 30 (2001) 9–18.
- [27] J.H. Patterson, R.F. Waller, D. Jeevarajah, H. Billman-Jacobe, M.J. McConville, Mannose metabolism is required for mycobacterial growth, *Biochem. J.* 372 (2003) 77–86.
- [28] H. Nankai, W. Hashimoto, K. Murata, Molecular identification of family 38 alpha-mannosidase of *Bacillus* sp strain GL1, responsible for complete depolymerization of xanthan, *Appl. Environ. Microbiol.* 68 (2002) 2731–2736.
- [29] M. Nakajima, H. Imamura, H. Shoun, T. Wakagi, Unique metal dependency of cytosolic alpha-mannosidase from *Thermotoga maritima*, a hyperthermophilic bacterium, *Arch. Biochem. Biophys.* 415 (2003) 87–93.
- [30] M.D.L. Suits, Y.P. Zhu, E.J. Taylor, J. Walton, D.L. Zechel, H.J. Gilbert, G.J. Davies, Structure and kinetic investigation of *Streptococcus pyogenes* family GH38 alpha-mannosidase, *PLoS One* 5 (2010).
- [31] A. Angelov, M. Putyrski, W. Liebl, Molecular and biochemical characterization of alpha-glucosidase and alpha-mannosidase and their clustered genes from the thermoacidophilic archaeon *Picrophilus torridus*, *J. Bacteriol.* 188 (2006) 7123–7131.
- [32] B. Cobucci-Ponzano, A. Trincone, A. Giordano, M. Rossi, M. Moracci, Identification of an archaeal alpha-L-fucosidase encoded by an interrupted gene – production of a functional enzyme by mutations mimicking programmed-1 frameshifting, *J. Biol. Chem.* 278 (2003) 14622–14631.
- [33] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* 72 (1976) 248–254.
- [34] R.J. Leatherbarrow, GraFit. Erithacus Software Ltd, Staines, U.K., 1992.
- [35] M.A. Frohman, Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs – thermal RACE, *Meth. Enzymol.* 218 (1993) 340–356.
- [36] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, *Bioinformatics* 23 (2007) 2947–2948.
- [37] D.H. Huson, D.C. Richter, C. Rausch, T. Dezulian, M. Franz, R. Rupp, Dendroscope: an interactive viewer for large phylogenetic trees, *BMC Bioinformatics* 8 (2007).
- [38] A. Bateman, E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S.R. Eddy, S. Griffiths-Jones, K.L. Howe, M. Marshall, E.L.L. Sonnhammer, The Pfam protein families database, *Nucleic Acids Res.* 30 (2002) 276–280.

- [39] S. Gribaldo, C. Brochier-Armanet, The origin and evolution of Archaea: a state of the art, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361 (2006) 1007–1022.
- [40] M. Moracci, B. Cobucci-Ponzano, A. Trincone, S. Fusco, M. De Rosa, J. van der Oost, C.W. Sensen, R.L. Charlebois, M. Rossi, Identification and molecular characterization of the first alpha-xylosidase from an Archaeon, *J. Biol. Chem.* 275 (2000) 22082–22089.
- [41] F.T. Zhao, J. Li, G.X. Shi, Y. Liu, L.P. Zhu, Modification of glycosylation reduces microvilli on rat liver epithelial cells, *Cell Biol. Int.* 26 (2002) 627–633.
- [42] E. Costanzi, C. Balducci, R. Cacan, S. Duvet, A. Orlacchio, T. Beccari, Cloning and expression of mouse cytosolic alpha-mannosidase (Man2cl), *Biochim. Biophys. Acta* 1760 (2006) 1580–1586.
- [43] E. Torarinsson, H.P. Klenk, R.A. Garrett, Divergent transcriptional and translational signals in Archaea, *Environ. Microbiol.* 7 (2005) 47–54.
- [44] M. Moracci, M. Ciaramella, M. Rossi, beta-Glycosidase from *Sulfolobus solfataricus*, *Meth. Enzymol.* 330 (2001) 201–215.
- [45] C. Notredame, D.G. Higgins, J. Heringa, T-Coffee: a novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol.* 302 (2000) 205–217.
- [46] H.D. Ly, S.G. Withers, Mutagenesis of glycosidases, *Annu. Rev. Biochem.* 68 (1999) 487–522.
- [47] S. Howard, C. Braun, J. McCarter, K.W. Moremen, Y.F. Liao, S.G. Withers, Human lysosomal and jack bean alpha-mannosidases are retaining glycosidases, *Biochem. Biophys. Res. Commun.* 238 (1997) 896–898.
- [48] E. Tarelli, H.L. Byers, M. Wilson, G. Roberts, K.A. Homer, D. Beighton, Detecting mannosidase activities using ribonuclease B and matrix-assisted laser desorption/ionization-time of flight mass spectrometry, *Anal. Biochem.* 282 (2000) 165–172.
- [49] K.K. Woo, M. Miyazaki, S. Hara, M. Kimura, Y. Kimura, Purification and characterization of a Co(II)-sensitive alpha-mannosidase from *Ginkgo biloba* seeds, *Biosci. Biotechnol. Biochem.* 68 (2004) 2547–2556.
- [50] D.R.P. Tulsiani, M.D. Skudlarek, S.K. Nagdas, M.C. Orgebinchrist, Purification and characterization of rat epididymal-fluid alpha-D-mannosidase – similarities to sperm plasma-membrane alpha-D-mannosidase, *Biochem. J.* 290 (1993) 427–436.
- [51] Y.Z. Jin, F. Dacheux, J.L. Dacheux, S. Bannai, Y. Sugita, N. Okamura, Purification and properties of major alpha-D-mannosidase in the luminal fluid of porcine epididymis, *Biochim. Biophys. Acta* 1432 (1999) 382–392.
- [52] E. Kuokkanen, W. Smith, M. Makinen, H. Tuominen, M. Puhka, E. Jokitalo, S. Duvet, T. Berg, P. Heikinheimo, Characterization and subcellular localization of human neutral class II alpha-mannosidase, *Glycobiology* 17 (2007) 1084–1093.
- [53] T. Yoshihisa, Y. Anraku, A novel pathway of import of alpha-mannosidase, a marker enzyme of vacuolar membrane, in *Saccharomyces cerevisiae*, *J. Biol. Chem.* 265 (1990) 22418–22425.
- [54] I. Benz, M.A. Schmidt, Never say never again: protein glycosylation in pathogenic bacteria, *Mol. Microbiol.* 45 (2002) 267–276.
- [55] B.N. Lillie, J.D. Hammermueller, J.I. MacInnes, M. Jacques, M.A. Hayes, Porcine mannan-binding lectin A binds to *Actinobacillus suis* and *Haemophilus parasuis*, *Dev. Comp. Immunol.* 30 (2006) 954–965.
- [56] J. Eichler, M.W.W. Adams, Posttranslational protein modification in Archaea, *Microbiol. Mol. Biol. Rev.* 69 (2005) 393.
- [57] M. Abu-Qarn, J. Eichler, N. Sharon, Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea, *Curr. Opin. Struct. Biol.* 18 (2008) 544–550.
- [58] B.H. Lower, P.J. Kennelly, The membrane-associated protein-serine/threonine kinase from *Sulfolobus solfataricus* is a glycoprotein, *J. Bacteriol.* 184 (2002) 2614–2619.
- [59] D.R. Francoleon, P. Boonthueung, Y. Yangt, U. Kim, A.J. Ytterberg, P.A. Denny, P.C. Denny, J.A. Loo, R.P. Gunsalus, R.R.O. Loo, S-layer, surface-accessible, and concanavalin A binding proteins of *Methanosarcina acetivorans* and *Methanosarcina mazei*, *J. Proteome Res.* 8 (2009) 1972–1982.
- [60] B. Zolghadr, A. Klingl, A. Koerd, A.J.M. Driessen, R. Rachel, S.-V. Albers, Appendage-mediated surface adherence of *Sulfolobus solfataricus*, *J. Bacteriol.* 192 (2010) 104–110.

# A New Archaeal $\beta$ -Glycosidase from *Sulfolobus solfataricus* SEEDING A NOVEL RETAINING $\beta$ -GLYCAN-SPECIFIC GLYCOSIDE HYDROLASE FAMILY ALONG WITH THE HUMAN NON-LYSOSOMAL GLUCOSYLCERAMIDASE GBA2\*

Received for publication, November 18, 2009, and in revised form, April 28, 2010. Published, JBC Papers in Press, April 28, 2010, DOI 10.1074/jbc.M109.086470

Beatrice Cobucci-Ponzano<sup>‡</sup>, Vincenzo Aurilia<sup>‡</sup>, Gennaro Riccio<sup>‡</sup>, Bernard Henrissat<sup>§</sup>, Pedro M. Coutinho<sup>§</sup>, Andrea Strazzulli<sup>‡</sup>, Anna Padula<sup>‡</sup>, Maria Michela Corsaro<sup>¶</sup>, Giuseppina Pieretti<sup>¶</sup>, Gabriella Pocsfalvi<sup>‡</sup>, Immacolata Fiume<sup>‡</sup>, Raffaele Cannio<sup>‡</sup>, Mosè Rossi<sup>‡||</sup>, and Marco Moracci<sup>‡1</sup>

From the <sup>‡</sup>Institute of Protein Biochemistry, CNR, Via P. Castellino 111, 80131, Naples, Italy, the <sup>§</sup>Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS, Universités Aix-Marseille I & II, 163 Avenue de Luminy, 13288 Marseille, France, and the <sup>¶</sup>Dipartimento di Chimica Organica e Biochimica and the <sup>||</sup>Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli Federico II, Complesso Universitario di Monte S. Angelo, Via Cinthia 4, 80126 Naples, Italy

Carbohydrate active enzymes (CAZymes) are a large class of enzymes, which build and breakdown the complex carbohydrates of the cell. On the basis of their amino acid sequences they are classified in families and clans that show conserved catalytic mechanism, structure, and active site residues, but may vary in substrate specificity. We report here the identification and the detailed molecular characterization of a novel glycoside hydrolase encoded from the gene *sso1353* of the hyperthermophilic archaeon *Sulfolobus solfataricus*. This enzyme hydrolyzes aryl  $\beta$ -gluco- and  $\beta$ -xylosides and the observation of transxylosylation reactions products demonstrates that SSO1353 operates via a retaining reaction mechanism. The catalytic nucleophile (Glu-335) was identified through trapping of the 2-deoxy-2-fluoroglucosyl enzyme intermediate and subsequent peptide mapping, while the general acid/base was identified as Asp-462 through detailed mechanistic analysis of a mutant at that position, including azide rescue experiments. SSO1353 has detectable homologs of unknown specificity among Archaea, Bacteria, and Eukarya and shows distant similarity to the non-lysosomal bile acid  $\beta$ -glucosidase GBA2 also known as glucocerebrosidase. On the basis of our findings we propose that SSO1353 and its homologs are classified in a new CAZy family, named GH116, which so far includes  $\beta$ -glucosidases (EC 3.2.1.21),  $\beta$ -xylosidases (EC 3.2.1.37), and glucocerebrosidases (EC 3.2.1.45) as known enzyme activities.

Carbohydrates, whose structural diversity exceeds by far the number of protein folds, are ubiquitous molecules that alone, or in form of glycoconjugates, mediate many biological processes (1). This extreme variety results from the diverse stereochemistry of the monosaccharide building blocks, from the enormous number of intersugar linkages they can form and to the fact that these molecules can decorate cell surface, large macromolecules (sugars themselves, proteins, nucleic acids), or small metabolites (lipids, antibiotics, etc). The breadth of the biological functions of carbohydrates, from the classical energetic and structural roles, is now well acknowledged although

the mechanisms of the sugar code are not known in detail. They include the control of the correct protein folding and activity (2), the mediation of molecular recognition events regulating cell-cell interactions (such as host-pathogen, cancer metastasis, etc.), cell signal transduction (nucleo-cytoplasm communication, differentiation, immune response, etc.) (for reviews see Refs. 1, 3, 4).

The ability of carbohydrates in functioning in intermolecular interactions as encoders of biological information is made possible by a large class of enzymes, collectively known as carbohydrate-active enzymes (CAZymes),<sup>2</sup> including glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases, carbohydrate esterases, and carbohydrate binding modules, which are in charge of catalyzing the metabolism and the correct shape of the sugars of the cell. CAZymes have been classified on the basis of their amino acid sequences in families sharing the same catalytic mechanism, structure, and active site residues; in addition, families with similar three-dimensional structure are further grouped in clans (5).

The number of CAZyme families is continuously increasing, thanks to the advent of new DNA sequencing techniques and subsequent sophisticated computer-aided sequence annotation procedures combined with new biochemical characterization. Interestingly, although the number of CAZy sequences increased 14-fold in the last 8 years, the number of enzymatic and structural characterization only doubled in the same time span, with at present <10% of total proteins in the CAZy database that have been characterized enzymatically (5). This contrast clearly shows that, in comparison with highly automated sequencing techniques, enzymatic characterization of novel

\* This work was supported by Project MoMa 1/014/06/0 of the Agenzia Spaziale Italiana.

<sup>1</sup> To whom correspondence should be addressed. Tel.: 390816132271; Fax: 390816132277; E-mail: m.moracci@ibp.cnr.it.

<sup>2</sup> The abbreviations used are: CAZyme, carbohydrate active enzyme; GBA1, lysosomal acid  $\beta$ -glucosidase; GH, glycoside hydrolases; GT, glycosyltransferases; ORF, open reading frame; GBA2, non-lysosomal glucosylceramidase; 4NP-Glc, 4-nitrophenyl- $\beta$ -D-glucopyranoside; 4NP-Xyl, 4-nitrophenyl- $\beta$ -D-xylopyranoside; MU-Glc, methylumbelliferyl- $\beta$ -D-glucopyranoside; MU-Xyl, methylumbelliferyl- $\beta$ -D-xylopyranoside; NB-DNJ, N-butyl-deoxynojirimycin; CBE, conduritol  $\beta$ -epoxide; TLC, thin layer chromatography; Ss $\beta$ -gly,  $\beta$ -glycosidase from *S. solfataricus*; 4NP-Xyl<sub>2</sub>, 4NP-disaccharide; 4NP-Xyl<sub>3</sub>, 4NP-trisaccharide; nano-ESI-MS/MS, nano-electrospray ionization tandem mass spectrometry; EPS, exo-polysaccharides; 2,4DNp-2F-Glc, 2,4-dinitrophenyl- $\beta$ -D-2-deoxy-2-fluoro-glucopyranoside; IDA, information-dependent acquisition; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight.



## Characterization of an Archaeal $\beta$ -Glycosidase

CAZymes is a longer and laborious process representing the limiting step for the full exploitation of genome sequencing efforts.

Here, we report the cloning, the heterologous expression and the detailed enzymatic characterization of a novel GH from the hyperthermophilic Archaeon *S. solfataricus*. This enzyme, encoded by the ORF SSO1353 displays sequence similarity to several unknown proteins from the three domains of life (Archaea, Bacteria, and Eukarya) and, more distantly, to human non-lysosomal glucosylceramidase, also known as  $\beta$ -glucosidase 2 (GBA2) (6), an enzyme involved in an alternative catabolic pathway of glucosylceramide (7).

The enzymatic characterization of the product of gene *sso1353* allowed us to demonstrate the retaining reaction mechanism followed by the enzyme, to evaluate its substrate specificity toward  $\beta$ -linked aromatic glucosides and xylosides, and to identify the catalytic amino acids in the active site. These results allowed us propose the role possibly played *in vivo* by this enzyme, which is expressed from a gene situated downstream of that coding for an endoglucanase in different *Sulfolobus* species. Finally, by virtue of the established commonalities within glycoside hydrolase families, the mechanistic data obtained with SSO1353 can be extended to all members of the newly created GH116 family including human GBA2.

### EXPERIMENTAL PROCEDURES

**Reagents**—All commercially available substrates were purchased from Sigma and Carbosynth. The GeneTailor Site-Directed Mutagenesis system was from Invitrogen, and the synthetic oligonucleotides were from PRIMM (Milan, Italy).

**Plasmid Preparation**—The SSO1353 ORF was cloned by amplification of *S. solfataricus*, strain P2, chromosomal DNA via PCR by using the following synthetic oligonucleotides: 1353Fw, 5'-ggaattccatggttacatatactgataagg-3', 1353Rv, 5'-tatcatgccatgctagaataggaagctcc-3', which introduce an NdeI and NcoI sites at the 5', just before the first ATG, and at the 3'-ends of the ORF, respectively. The program was as follows: 5 min at 95 °C, 1.5 min at 50 °C, and 4 min at 72 °C; 30 cycles at 95 °C for 45 s, 50 °C for 1.5 min, and 72 °C for 4 min; final extension at 72 °C for 10 min. The resulting DNA fragment was cloned in the pET29a plasmid (Novagen), obtaining the vector pET1353, in which the SSO1353 ORF is under the control of the isopropyl-1-thio- $\beta$ -D-galactopyranoside inducible T7 RNA polymerase promoter that drives high expression levels in bacterial hosts. The ORF obtained after amplification was controlled by DNA sequencing.

**Site-directed Mutagenesis**—The mutants E335G, D406G, D458G, D462G were prepared by site-directed mutagenesis from the pET1353 plasmid, by following the instructions of the manufacturer. The mutagenic oligonucleotides were the following (mismatches are underlined): E335Gmut, 5'-ATGG-ACAGTTCTGTGGTGGCTCCGTAAATCGCA-3'; E335Grev, 5'-AGCACCACAGAACTGTCCATATTTAGGTAC-3'; D406Gmut, 5'-TTCTCCTCCCAGATGGAAGGGTATGAA-TCCGA-3'; D406Grev, 5'-CCTTCCATCTGGGAGGAGAA-GTAGTACCAT-3'; D458Gmut, 5'-ATTCATGGAGGGAG-AGATGGGCAATGCTTTTG-3'; D458Grev, 5'-CCATCTC-TCCCTCCATGAATGGTAAATTCC-3'; D462Gmut, 5'-

AGAGATGGACAATGCTTTTGGCGCTACCATCA-3'; D462Grev, 5'-CAAAAGCATTGTCCATCTCTCCCTCCA-TGA-3'. The genes containing the desired mutation were identified by direct sequencing and completely resequenced.

**Expression and Purification of SSO1353 Wild Type and Mutants**—*Escherichia coli* BL21(DE3)Ril/pET1353 wild type and mutants were grown in 2 liters of LB at 37 °C supplemented with kanamycin (50  $\mu$ g/ml) and chloramphenicol (30  $\mu$ g/ml). Gene expression was induced by the addition of 0.5 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside when the culture reached an  $A_{600}$  of 1.0. Growth was allowed to proceed for 16 h, and cells were harvested by centrifugation at  $5,000 \times g$ . The resulting cell pellet was thawed, resuspended in 3 ml  $g^{-1}$  cells of 20 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, 1% (v/v) Triton X-100 and homogenized by French cell pressure treatment. After centrifugation for 30 min at  $10,000 \times g$ , the crude extract was incubated with Benzonase (Novagen) for 1 h at room temperature and then heat-fractionated for 30 min at 55 and 75 °C and for 20 min at 85 °C. The supernatant obtained after heat fractionations, equilibrated in 1 M ammonium sulfate, was applied to a HiLoad 26/10 phenyl Sepharose high performance (Amersham Biosciences), which had been equilibrated with 20 mM sodium phosphate buffer, pH 7.3, 1 M ammonium sulfate. After washing with 2 column volumes with the loading buffer, the protein was eluted with a linear gradient of water at a flow rate of 3 ml  $min^{-1}$ ; the protein eluted in 100% water. Active fractions were pooled, equilibrated in 20 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl and concentrated by ultrafiltration on an Amicon YM30 membrane (cut off 30,000 Da). For the wild-type enzyme, after concentration, the sample was loaded onto a HiLoad 26/60 Superdex 200 prep grade column (Amersham Biosciences). Active fractions were pooled and concentrated; protein concentration was determined with the method of Bradford (8). The SSO1353 wild type and mutants were 95% pure by SDS-PAGE and were stored at 4 °C.

**Characterization of SSO1353 Wild Type and Mutants**—The molecular mass of native SSO1353 wild type was determined by gel filtration on a Superdex 200 HR 10/30 FPLC column (Amersham Biosciences); molecular weight markers were albumin (66,000), alcohol dehydrogenase (150,000),  $\beta$ -amylase (200,000), and apoferritin (443,000).

The standard assay for SSO1353 activity was performed in 50 mM sodium citrate buffer at pH 5.5 at 65 °C on the indicated substrates. Typically, in each assay we used 1–10  $\mu$ g of SSO1353 in the final volume of 1.0 ml. Kinetic constants of SSO1353 wild type and D462G mutant on aryl glycosides 4Np-Glc, 4Np-Xyl, and 2Np-Glc were measured at standard conditions at 65 °C by using concentrations of substrate ranging between 1 and 150 mM. The  $\epsilon_{m,M}$  extinction coefficients at 405 nm for 2- and 4-nitrophenol under standard conditions and 65 °C were 1.1 and 3.3  $mm^{-1} cm^{-1}$ , respectively. One unit of enzyme activity was defined as the amount of enzyme catalyzing the hydrolysis of 1  $\mu$ mol of substrate in 1 min at the conditions described. The metal dependence of the wild-type enzyme was evaluated using 5 mM 4Np-Xyl as substrate in the presence of 1 mM EDTA and 5 mM  $MgCl_2$  or  $MnCl_2$ , in standard conditions.

The activity of the wild-type enzyme on different aryl glycosides was tested in 50 mM sodium citrate buffer at pH 5.5 at 65 °C. Typically, in each assay we used 70  $\mu$ g of SSO1353 in a final volume of 0.2 ml. The reaction was started by adding the enzyme, and it was stopped by adding 0.8 ml of 1 M iced sodium carbonate. The optical density of the solution was measured at 420 nm at room temperature. The molar extinction coefficients of 4-nitrophenol and 2-nitrophenol, measured at 420 nm, at room temperature and in 1 M sodium carbonate buffer were 17.2 and 4.7  $\text{mm}^{-1} \text{cm}^{-1}$ , respectively. In all of the assays, spontaneous hydrolysis of the substrate was subtracted by using appropriate blank mixtures without the enzyme.

The chemically rescued activity of the D462G mutant was measured in 50 mM sodium citrate buffer at pH 5.5 on 40 mM 2Np-Glc at 65 °C as described above; where indicated, the assay mixture was supplemented with the indicated concentrations of sodium azide as external nucleophile. In all of the assays, spontaneous hydrolysis of the substrate was subtracted by using appropriate blank mixtures without the enzyme. Aliquots of the reaction mixtures were analyzed on a silica gel 60 F<sub>254</sub> TLC by using ethyl acetate/methanol/water (70:20:10 v/v) as eluant and were detected by exposure to 4%  $\alpha$ -naphthol in 10% sulfuric acid in ethanol followed by charring. The  $\beta$ -D-glucosyl azide isolated from the enzymatic reaction mixture of the D462G mutant was identified by <sup>1</sup>H and <sup>13</sup>C NMR spectroscopy. NMR:  $\delta$  4.75 H-1 (d, <sup>3</sup>J<sub>H1,H2</sub> = 8.8 Hz),  $\delta$  91.4 C-1;  $\delta$  3.25 H-2 (t),  $\delta$  74.0 C-2;  $\delta$  3.51 H-3 (t),  $\delta$  76.9 C-3;  $\delta$  3.41 H-4 (t),  $\delta$  70.5 C-4;  $\delta$  3.53 H-5 (m),  $\delta$  79.0 C-5;  $\delta$  3.73 H-6<sub>a</sub> (dd),  $\delta$  3.90 H-6<sub>b</sub> (dd), 61.6 C-6.

The activity of the wild-type enzyme in the presence of 0.5–2.0 mg ml<sup>-1</sup> Triton X-100 or CHAPS was determined on 20 mM 4Np-Glc as described above. The activity of the wild type on  $\beta$ -D-oligosaccharides of glucose (G<sub>3</sub> to G<sub>5</sub>) (5 mM) and xylose (X<sub>2</sub> to X<sub>5</sub>) (2.5 mM) was tested in 50 mM sodium citrate buffer at pH 5.5, by using 12–63  $\mu$ g of enzyme, at 65 °C in a final volume of 0.2 ml. Aliquots of the reaction mixtures were analyzed by TLC by using ethyl acetate/acetic acid/isopropanol/formic acid/water (50:20:10:2:30 v/v) (for G<sub>3</sub>–G<sub>5</sub>) or acetone/isopropyl alcohol/water (60:30:15 v/v) (for X<sub>2</sub>–X<sub>5</sub>) as the eluant and were detected as described above. The activity of the wild type on glucocerebroside (Matreya), octyl- $\beta$ -D-glucopyranoside (Sigma), and gangliosides, measured at the same conditions was analyzed by TLC by using chloroform/methanol/CaCl<sub>2</sub> 15 mM (60:40:9 v/v).

The steady-state kinetic constants of the wild type on MU-Glc and MU-Xyl were calculated by following a slight modified method already described (9). Briefly, the enzyme activity was determined fluorimetrically with MU-Glc (0.1–6 mM) and MU-Xyl (0.025–5 mM) as substrates in 50 mM sodium citrate buffer at pH 5.5, by using 1  $\mu$ g of the enzyme. Assays (0.25 ml final volume) were conducted for 1–5 min at 65 °C, and the reaction was stopped with 0.5 ml of 0.1 M glycine-NaOH buffer, pH 10.3. The formation of the methylumbelliferone was measured by emission at 450 nm with excitation at 384 nm. In all of the assays, spontaneous hydrolysis of the substrate was subtracted by using appropriate blank mixtures without the enzyme. All kinetic data were calculated as the average of at

least two experiments and were plotted and refined with the program GraFit (10).

**Inhibition of SSO1353 Wild Type**—The effect of the inhibitor 2,4-dinitrophenyl- $\beta$ -D-2-deoxy-2-fluoro-glucopyranoside (2,4DNp-2F-Glc) (Sigma) was analyzed as previously performed by Shaikh *et al.* (11). Briefly, wild-type SSO1353 (0.1  $\mu$ g/ $\mu$ l) was incubated at 45 °C in mixtures containing 0.3, 3.0, 7.0, and 18.0 mM concentrations of inhibitor and 50 mM sodium citrate buffer, pH 5.5. An identical mixture containing all the reagents with the exception of the enzyme was prepared as control. At time intervals, aliquots from the two mixtures were withdrawn and used to measure the enzymatic activity and as blank, respectively. Assays were performed on 60 mM 2Np-Glc in standard conditions. Initial rates at each time point were elaborated as described in Ref. 11 to measure the inactivation parameters  $K_i$  and  $k_i$  with GraFit (10).

To determine the effect of the inhibitors *N*-butyldeoxynojirimycin (NB-DNJ) and conduritol  $\beta$ -epoxide (CBE), SSO1353 (11  $\mu$ g) was incubated in the presence of increasing concentrations of the inhibitors (0.1–5 mM for NB-DNJ and 0.1–2 mM for CBE) in 50 mM sodium citrate buffer, pH 5.5, for 30 min at 45 °C, in a final volume of 80  $\mu$ l. Identical mixtures containing all the reagents with the exception of the inhibitor were used to determine the 100% of activity. After incubation, the samples were diluted 3-fold in 5 mM MU-Glc, 50 mM sodium citrate buffer, pH 5.5 in a final volume of 0.25 ml, and assayed as described above.

**Transglycosylation Reactions**—Wild-type SSO1353 (7–148  $\mu$ g), was incubated in 50 mM sodium citrate buffer pH 5.5 with 5 mM 4Np-Xyl or 4Np-Glc, or both, at 5 mM each, at 65 °C for 16 h in a final volume of 0.2–1 ml. Identical mixtures containing all the reagents but the enzyme were prepared as control. Where specified, after incubation, 50  $\mu$ g of the  $\beta$ -glycosidase from *S. solfataricus* (Ss $\beta$ -gly) were added to aliquots (90  $\mu$ l) of the reaction mixtures and further incubated 2 h at 65 °C. All the reactions were examined by TLC by using ethyl acetate/methanol/water (70:20:10 v/v) as eluant as described above.

**Identification of the Reaction Products**—The <sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded in D<sub>2</sub>O at 600 MHz with a spectrometer equipped with a cryo probe, in the FT mode at 303 K. <sup>1</sup>H chemical shifts are expressed in  $\delta$  relative to HOD signal (4.72 ppm).

The linkage analysis of the oligosaccharide was obtained by methylation according to Ciucanu procedure, as already reported (12). The sample obtained was injected into GC-MS and partially methylated alditol acetates were recognized from their EI-MS spectra and by comparison with pure synthetic standards. Partially methylated alditol acetates were analyzed on an Agilent Technologies gas chromatograph 6850A equipped with a mass selective detector 5973N and a Zebron ZB-5 capillary column (Phenomenex, 30 m  $\times$  0.25 mm i.d., flow rate 1 ml/min, He as carrier gas). The temperature program was: 90 °C for 1 min, 90 °C  $\rightarrow$  140 °C at 25 °C min<sup>-1</sup>, 140 °C  $\rightarrow$  200 °C at 5 °C min<sup>-1</sup>, 200 °C  $\rightarrow$  280 °C at 10 °C min<sup>-1</sup>, 280 °C for 10 min.

The reaction mixture was purified by reverse phase chromatography (Polar-RP 80A, Phenomenex, 4  $\mu$ , 250  $\times$  10 mm) on an Agilent HPLC instrument 1100 series, using H<sub>2</sub>O/CH<sub>3</sub>OH

## Characterization of an Archaeal $\beta$ -Glycosidase

6/4 with 20 mM trifluoroacetic acid (final concentration) as eluant. The eluted products were first analyzed by positive ions reflectron MALDI-TOF mass spectrometry.

The trisaccharide  $\beta$ -Xyl-(1 $\rightarrow$ 4)- $\beta$ -Xyl-O-4-Np showed a pseudomolecular ion ( $M+Na$ )<sup>+</sup> at  $m/z$  425.94 (calculated  $m/z$  426.10) and the methylation analysis indicated the presence of terminal xylopyranose and 4-substituted xylopyranose. NMR:  $\delta$  5.24 H-1<sub>A</sub> (d,  $J_{H-1,H-2}$  = 7.3 Hz),  $\delta$  101.0 C-1<sub>A</sub>; 3.66 H-2<sub>A</sub> (t),  $\delta$  73.7 C-2<sub>A</sub>;  $\delta$  3.72 H-3<sub>A</sub> (t),  $\delta$  74.5 C-3<sub>A</sub>;  $\delta$  3.89 H-4<sub>A</sub> (m),  $\delta$  77.4 C-4<sub>A</sub>;  $\delta$  4.19 H<sub>a</sub>-5<sub>A</sub> (dd),  $\delta$  3.62 H<sub>b</sub>-5<sub>A</sub> (t),  $\delta$  64.3 C-5<sub>A</sub>;  $\delta$  4.49 H-1<sub>B</sub> (d,  $J_{H-1,H-2}$  = 7.3 Hz),  $\delta$  103.1 C-1<sub>B</sub>; 3.29 H-2<sub>B</sub> (t),  $\delta$  74.0 C-2<sub>B</sub>;  $\delta$  3.44 H-3<sub>B</sub> (t),  $\delta$  76.9 C-3<sub>B</sub>;  $\delta$  3.62 H-4<sub>B</sub> (m),  $\delta$  70.4 C-4<sub>B</sub>;  $\delta$  3.99 H<sub>a</sub>-5<sub>B</sub> (dd),  $\delta$  3.32 H<sub>b</sub>-5<sub>B</sub> (t),  $\delta$  66.5 C-5<sub>B</sub>.

The trisaccharide  $\beta$ -Xyl-(1 $\rightarrow$ 4)- $\beta$ -Xyl-(1 $\rightarrow$ 4)- $\beta$ -Xyl-O-4-Np showed a pseudomolecular ion in a MALDI-TOF-MS ( $M+Na$ )<sup>+</sup> at  $m/z$  557.82 (calculated  $m/z$  558.15); methylation analysis: terminal xylopyranose and 4-substituted xylopyranose.

NMR:  $\delta$  5.25 H-1<sub>A</sub> (d,  $J_{H-1,H-2}$  = 7.3 Hz),  $\delta$  101.2 C-1<sub>A</sub>; 3.66 H-2<sub>A</sub> (t),  $\delta$  73.8 C-2<sub>A</sub>;  $\delta$  3.72 H-3<sub>A</sub> (t),  $\delta$  74.6 C-3<sub>A</sub>;  $\delta$  3.90 H-4<sub>A</sub> (m),  $\delta$  77.3 C-4<sub>A</sub>;  $\delta$  4.19 H<sub>a</sub>-5<sub>A</sub> (dd),  $\delta$  3.62 H<sub>b</sub>-5<sub>A</sub> (t),  $\delta$  64.3 C-5<sub>A</sub>;  $\delta$  4.51 H-1<sub>B</sub> (d,  $J_{H-1,H-2}$  = 7.3 Hz),  $\delta$  103.0 C-1<sub>B</sub>; 3.31 H-2<sub>B</sub> (t),  $\delta$  74.0 C-2<sub>B</sub>;  $\delta$  3.56 H-3<sub>B</sub> (t),  $\delta$  75.0 C-3<sub>B</sub>;  $\delta$  3.80 H-4<sub>B</sub> (m),  $\delta$  77.6 C-4<sub>B</sub>;  $\delta$  4.12 H<sub>a</sub>-5<sub>B</sub> (dd),  $\delta$  3.39 H<sub>b</sub>-5<sub>B</sub> (t);  $\delta$  64.3 C-5<sub>B</sub>;  $\delta$  4.47 H-1<sub>C</sub> (d,  $J_{H-1,H-2}$  = 7.3 Hz),  $\delta$  103.1 C-1<sub>C</sub>;  $\delta$  3.26 H-2<sub>C</sub> (t),  $\delta$  74.1 C-2<sub>C</sub>;  $\delta$  3.43 H-3<sub>C</sub> (t),  $\delta$  76.9 C-3<sub>C</sub>;  $\delta$  3.63 H-4<sub>C</sub> (m),  $\delta$  70.4 C-4<sub>C</sub>;  $\delta$  3.97 H<sub>a</sub>-5<sub>C</sub> (dd),  $\delta$  3.31 H<sub>b</sub>-5<sub>C</sub> (t);  $\delta$  66.5 C-5<sub>C</sub>.

**Nano-ESI-MS of Intact Protein Samples**—Samples were analyzed using a triple quadrupole time of flight instrument (QSTAR Elite, Applied Biosystems, Foster City, CA/Toronto, Canada) equipped with a nanoflow electrospray ion source. Pulled silica capillary (170  $\mu$ m outer diameter/100  $\mu$ m inner diameter, tip 30  $\mu$ m inner diameter) was used as nanoflow tip. For the analysis of intact proteins, 4  $\mu$ g of samples were purified using ZipTip C4 (Millipore, Billerica, MA). Proteins were eluted by 50% acetonitrile and 0.1% formic acid. Purified proteins (10  $\mu$ M) were loaded into the ion source at 300 nl/min flow rate using a syringe pump. Single-stage ESI mass spectra were acquired in the range of  $m/z$  300–2000. For protein molecular mass determination three independent measurements were performed. The expected mass error on the average molecular mass of intact proteins was about  $\pm 0.01\%$ . For data acquisition and Bayesian protein reconstruction the *Analyst QS 2.0* software (Applied Biosystems, Foster City, CA/Toronto, Canada) was used.

**Nano-HPLC-ESI-MS/MS Experiments**—Wild-type SSO1353 (22  $\mu$ g, 0.3 nmol) was incubated with 2.9 mM 2,4-dinitrophenyl- $\beta$ -D-2-deoxy-2-fluoro-glucopyranoside (2,4DNP-2F-Glc) (Sigma) at 1:1000 enzyme/inhibitor ratio in 50 mM sodium citrate buffer, pH 5.5 at 45 °C. An identical mixture containing all the reagents with the exception of the inhibitor was prepared as control. At time intervals, aliquots from the two mixtures were withdrawn and assayed on 60 mM 2Np-Glc in standard conditions.

Samples (0.086  $\mu$ g/ $\mu$ l) were enzymatically digested in the acidified inhibition buffer (formic acid to 5% (v/v) final concentration, pH 2) using pepsin from porcine stomach mucosa (3,260 units/mg, Sigma-Aldrich) at 1:20 enzyme to substrate

ratio at 37 °C for 30 min. Resulting peptide mixtures (5  $\mu$ l) were loaded, purified, and concentrated on a monolithic trap column (200  $\mu$ m inner diameter  $\times$  5 mm, LCPackings, Sunnyvale, CA) at 25  $\mu$ l/min flow rate and separated by nanoflow reverse-phase chromatography on a PS-DVB monolithic column (200  $\mu$ m inner diameter  $\times$  5 cm, LCPackings) at 300 nl/min using an UltiMate<sup>TM</sup> 3000 HPLC (Dionex, Sunnyvale, CA). The following solvents and gradient conditions were used: solvent A: 2% acetonitrile in 0.1% formic acid and 0.025% trifluoroacetic acid, solvent B: 98% acetonitrile in 0.1% formic acid and 0.025% trifluoroacetic acid, gradient: 5–50% B in 40 min, 50–98% B in 6 s. Eluting peptides were directly analyzed by nano-ESI-MS in positive ion mode using information-dependent acquisition (IDA). The two most abundant multiply charged ions were automatically selected and subjected for collision induced dissociation experiments. Nitrogen was used as collision gas. Tandem mass spectra were analyzed by manual inspection and by the use of Mascot Server (version 2.2). Peak lists for Mascot containing all acquired MS/MS spectra were generated by Analyst QS 2.0 software using the default parameters. Mascot was set up to search database containing a single protein (SSO1353) sequence extracted from NCBI and was run with a fragment ion mass tolerance of 0.1 Da and a parent ion tolerance of 50 ppm. MS/MS ion score cut-off was set to 10. No enzyme was specified. 2F-Glc was defined as variable modification in Mascot searches. Three independent inhibition experiments were performed and on each resulting samples two analytical measurements were run.

**Definition and Analysis of a New Glycoside Hydrolase Family**—Gapped BLAST searches (13) were performed against the non-redundant protein set at the NCBI and against classified and unclassified sequences present in carbohydrate-active enzymes database (CAZy) (5). A total of 90 sequences were used to define new family. This family includes proteins from archaeal, bacterial, and eukaryotic origin, several already collected in CAZy by similarity to human non-lysosomal bile acid  $\beta$ -glucosidase 2 (Gba2), but not yet assigned to a family. This family was designated as glycoside hydrolase family 116 and will be released in CAZy. The sequences were aligned with Muscle 3.7 (14), and the resulting alignments were subsequently manipulated and analyzed with an in-house modified version of Jalview (15).<sup>3</sup> Estimated sequence distances were determined by maximum likelihood using LG distances (16) and constructed a distance tree using the Ward hierarchical clustering method (17).

## RESULTS

**Isolation of ORF SSO1353**—The inspection of the genomic sequence of the archaeon *S. solfataricus*, strain P2, revealed an ORF downstream of the gene *sso1354* encoding for an endoglucanase (Fig. S1) (18). *Sso1353* is presently annotated as an hypothetical protein while the other ORFs in this cluster, *sso1351*, *sso1352*, and *sso1355*, are a putative permease, a transcriptional regulator, and a carboxypeptidase, respectively. ORFs *sso1354* and *sso1353* are transcribed in the same direction and are separated by 57 bp in which the latter ORF is preceded by a putative

<sup>3</sup> P. M. Coutinho and B. Henrissat, unpublished data.



promoter formed by an AT-rich box A (centered at  $-30$  nt from ATG) and a TFB-responsive element (centered at  $-38$  nt) (not shown). Northern blot analysis showed that *ssol1353* is expressed as an isolated gene (not shown) and the absence of a clear Shine-Dalgarno-like motif in the intergenic region suggests that *ssol1353* gene is translated as a leaderless gene (19). Initial gapped BLAST searches (13) revealed that SSO1353 is similar to proteins of unknown function from Archaea, Bacteria, and Eukarya and, to a lesser extent, to eukaryotic non-lysosomal bile acid  $\beta$ -glucosidases. The higher sequence identity scores ( $>31\%$ ) were with archaeal proteins with the highest (86%) with loci *ssol1948* from *S. solfataricus*, strain P2, and M1425\_0924 and M1627\_099 from *S. islandicus*, strains M.14.25 and L.S.2.15, respectively. Interestingly, these highly similar genes also lie downstream to a locus encoding for an endoglucanase (SSO1949 in *S. solfataricus* (20)), suggesting that gene duplication occurred in these organisms. Among non-lysosomal bile acid  $\beta$ -glucosidases, the scores were much lower, the best ones being with human and *Ciona intestinalis* enzymes (19% identity). Sequences from the NCBI were searched to complement the set of unclassified glycoside hydrolase sequences already present in CAZy previously collected based on the bile acid-glycosidases, and integrated to create a new family, that we have designated as Glycoside Hydrolase family 116 (GH116). Once the conserved catalytic regions were aligned, a distances tree was obtained (Fig. 1).

To ascertain if the *ssol1353* encodes for a novel glycoside hydrolase, the corresponding gene was cloned by PCR from the genomic DNA of *S. solfataricus*, strain P2. The primer at the 5' of the gene was designed starting from the first Met. Attempts to express SSO1353 fused to glutathione *S*-transferase (at the N terminus) or to a His tag (at the C terminus) were unsuccessful; therefore, the gene was cloned in pET29a without any purification tag, obtaining the plasmid vector pET1353. The resulting recombinant SSO1353 protein was successfully expressed in the soluble fraction and purified to homogeneity by performing three subsequent heating steps followed by a hydrophobic chromatography and a gel filtration (Fig. 2). After the last purification step we obtained about 1.5 mg of pure protein per liter of *E. coli* culture.

**Properties of SSO1353**—A gel filtration run with the suitable molecular weight standards revealed that SSO1353 was a monomer of about 76 kDa in native conditions (not shown). The purified enzyme was optimally active at pH 5.5 (50 mM sodium citrate) when assayed on 5 mM 4-nitrophenyl- $\beta$ -D-glucopyranoside (4Np-Glc, 4Np-Xyl) substrates at 65 °C, with specific activities of 0.3 and 0.8 units  $\text{mg}^{-1}$ , respectively. The activity on 4Np-Xyl was not affected by Triton X-100 (0.5–2.0  $\text{mg ml}^{-1}$ ), CHAPS (0.5–2.0  $\text{mg ml}^{-1}$ ),  $\text{Mn}^{2+}$ ,  $\text{Mg}^{2+}$  (5 mM), and 1 mM EDTA.

To determine the substrate specificity of this enzyme, the hydrolytic activity of SSO1353 toward various substrates was investigated at 65 °C in 50 mM sodium citrate, pH 5.5. In addition to 4Np-Glc and -Xyl, also 4Np-Gal, 2Np-Glc, -Gal, -Xyl, methylumbelliferyl- $\beta$ -D-glucopyranoside (MU-Glc), and MU-Xyl were substrates of the enzyme. Instead, no activity was observed on 4Np-Man and -GlcNAc; X-Glc and X-Gal; 2Np-cellobioside; 4Np- $\alpha$ -D-Gal-Glc, and, -Man; 4Np- $\alpha$ -L-Fuc, -ar-

abinoside; 4Np- $\beta$ -L-Fuc,  $\beta$ -D-oligosaccharides of glucose and xylose (di-, tri-, tetra- and pentaose), and glucocerebroside, octyl- $\beta$ -D-glucopyranoside and gangliosides even after prolonged incubations. Steady-state kinetic parameters of SSO1353 were measured for the glycosides that were substrates of the enzyme (Table 1). SSO1353 showed similar kinetic constants for the substrates tested: the highest specificity constant for MU-glycosides results from a reduced  $K_m$ . Therefore, this kinetic characterization showed that SSO1353 is a  $\beta$ -D-glycosidase specific for gluco- and xylosides (EC 3.2.1.21/37) showing increased affinity for substrates having hydrophobic leaving groups. Finally, interestingly, the activity of SSO1353 on MU-Glc was inhibited by both *N*-butyl-deoxynojirimycin (NB-DNJ) and conduritol  $\beta$ -epoxide (CBE). The former showed an  $\text{IC}_{50}$  of 1.5 mM while 2 mM CBE gave 93% inhibition after 30 min of incubation. The sensitivity to this irreversible inhibitor differentiates SSO1353 from human non-lysosomal glucosylceramidase, which is insensitive to CBE (see below) (6).

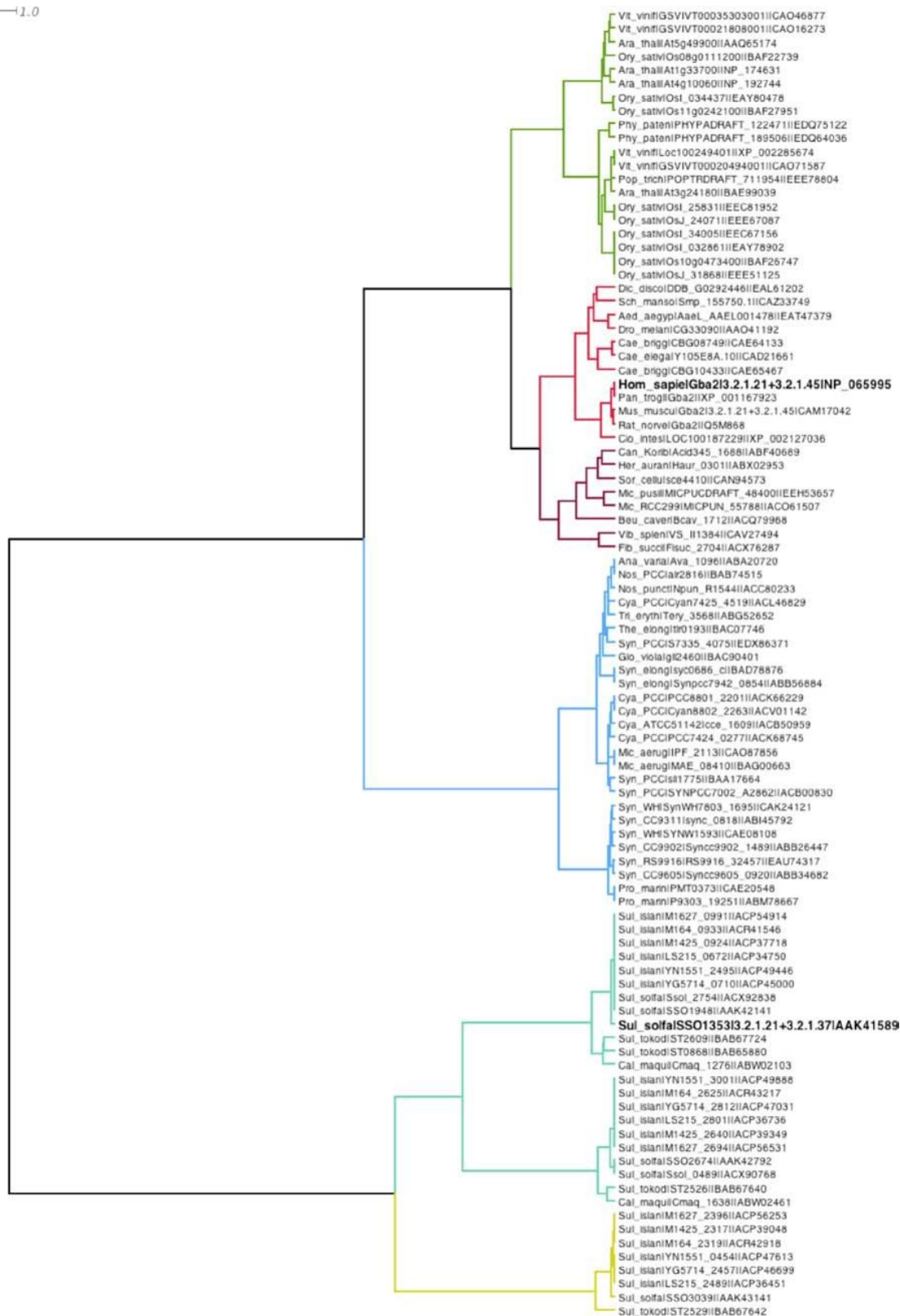
**SSO1353 Has Transglycosylation Activity and Follows a Retaining Reaction Mechanism**—The products of the reaction mixtures containing SSO1353 and 4Np-Xyl were examined by thin layer chromatography (TLC) (Fig. 3). Interestingly, the products include not only xylose, but also oligosaccharides with a higher degree of polymerization than the 4Np-Xyl substrate, indicating that the enzyme performed transglycosylation reactions (Fig. 3A, lane 4). The transglycosylation products could be completely hydrolyzed (Fig. 3A, lane 5) by the addition of limiting amounts of  $\beta$ -glycosidase from *S. solfataricus* (Ss $\beta$ -gly), which shows broad substrate specificity for  $\beta$ -D-glycosides (22), indicating that SSO1353 catalyzed the formation of  $\beta$ -D-xylo-oligosaccharides. Remarkably, the enzyme catalyzed transglycosylation reactions also by using 4Np-Glc (5 mM) as a substrate and an increased number of transglycosylation products were found when both substrates were included in the reaction mixture: at least five compounds are easily observable by TLC (Fig. 3B, lane 8).

To determine the stereo- and regioselectivity of the transglycosylation activity of SSO1353 in the presence of 4Np-Xyl we scaled up the reaction to purify the products. The reverse phase HPLC purification revealed the presence of two transglycosylation products corresponding to 4NP-disaccharide (4NP-Xyl<sub>2</sub>) and 4NP-trisaccharide (4NP-Xyl<sub>3</sub>). Each product was identified by MALDI-TOF mass spectrometry, methylation analysis,  $^1\text{H}$  and  $^{13}\text{C}$ -NMR spectroscopy. The positive ions MALDI-TOF mass spectrum of 4NP-Xyl<sub>2</sub> showed a pseudomolecular ion  $(\text{M}+\text{Na})^+$  at  $m/z$  425.84, which accounted for the presence of a 4-phenyl glycoside of xylose disaccharide.

The  $^1\text{H}$ -NMR spectrum showed two anomeric doublets in the ratio of 1:1 of xylose at 5.24 ppm (H-1 of  $\beta$  anomer A,  $^{1,2}J_{\text{H,H}}$  7.3 Hz) and 4.49 ppm (H-1 of  $\beta$  anomer B,  $^{1,2}J_{\text{H,H}}$  7.3 Hz), respectively, together with signals of protons geminal to hydroxyl groups in the region between 3.0 and 4.0 ppm. Moreover signals having the same intensity as anomeric signals and attributable to the 4-nitrophenyl moiety, occurred at 7.25, and 8.28 ppm. The chemical shifts of these signals were in agreement with those of the disaccharide  $\beta$ -Xyl-(1 $\rightarrow$ 4)- $\beta$ -Xyl-O-4-Np, as obtained by analysis of two-dimensional  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy. In particular the (1 $\rightarrow$ 4) linkage was

# Characterization of an Archaeal $\beta$ -Glycosidase

1.0



Downloaded from www.jbc.org at ADRIANO BUZZATI-TRAVERSO, on November 17, 2010

## Characterization of an Archaeal $\beta$ -Glycosidase

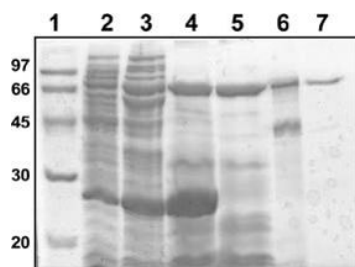


FIGURE 2. **SDS-PAGE analysis of SSO1353.** Lane 1, molecular weight markers; lane 2, *E. coli* BL21(DE3)/Ril/pET1353 soluble protein extract (45  $\mu$ g); lanes 3–5, protein extract after heat treatment at 55, 75, and 85  $^{\circ}$ C, respectively (70, 120 and 84  $\mu$ g, respectively); lane 6, typical sample after hydrophobic chromatography (21  $\mu$ g); lane 7, sample after gel filtration (3  $\mu$ g).

**TABLE 1**  
Steady-state kinetic constants of SSO1353

Substrate	$k_{cat}$ $s^{-1}$	$K_m$ mM	$k_{cat}/K_m$ $s^{-1} mM^{-1}$
2Np-Glc	$4.7 \pm 0.3$	$13 \pm 2$	0.37
4Np-Glc	$4.9 \pm 0.5$	$54 \pm 12$	0.09
4Np-Xyl	$4.3 \pm 0.4$	$25 \pm 8$	0.17
MU-Glc	$1.2 \pm 0.1$	$2.6 \pm 0.5$	0.47
MU-Xyl	$0.8 \pm 0.1$	$1.2 \pm 0.3$	0.71

deduced from the C-4 glycosylation shift at 77.4 ppm of xylose unit A respect to the value of 70.4 ppm for an unsubstituted xylopyranoside (23). Further support to this structure derived from the methylation analysis, which showed the presence of 1,4,5-tri-*O*-acetyl-2,3-di-*O*-methyl xylitol, corresponding to 4-substituted xylopyranose unit, and of 1,5-di-*O*-acetyl-2,3,4-tri-*O*-methyl xylitol, corresponding to a terminal non-reducing end xylopyranose unit.

The positive ions MALDI-TOF mass spectrum of 4NP-Xyl<sub>3</sub> showed a pseudomolecular ion ( $M + Na$ )<sup>+</sup> at  $m/z$  557.52, which accounted for the presence of a nitro-phenyl glycoside of xylose trisaccharide. The <sup>1</sup>H NMR spectrum showed the presence of three anomeric proton signals at 5.25, 4.51, and 4.47 ppm, respectively. The beta anomeric configuration for all the xylose units was inferred from the 7.3 Hz value of <sup>1,2</sup>*J*<sub>H,H</sub> coupling constants. The methylation analysis revealed the same residues as for 4NP-Xyl<sub>2</sub>, suggesting a linear trisaccharide structure  $\beta$ -Xyl-(1 $\rightarrow$ 4)- $\beta$ -Xyl-(1 $\rightarrow$ 4)- $\beta$ -Xyl-*O*-4-Np for 4NP-Xyl<sub>3</sub>. The complete assignment of the <sup>1</sup>H and <sup>13</sup>C values (see "Experimental Procedures") confirmed the above structure. These results unequivocally demonstrate that SSO1353 promoted the transglycosylation reaction by following a retaining reaction mechanism.

**Identification of the Catalytic Residues of SSO1353 by Site-directed Mutagenesis**—Retaining glycosidases generally utilize a double displacement mechanism catalyzed by two enzymatic carboxylates and in which a glycosyl intermediate is formed and hydrolyzed. In the first step of the reaction, one of the carboxylic acids functions as a general acid catalyst protonating the glycosidic oxygen while the nucleophile residue attacks the sugar anomeric center to form the glycosyl enzyme intermedi-

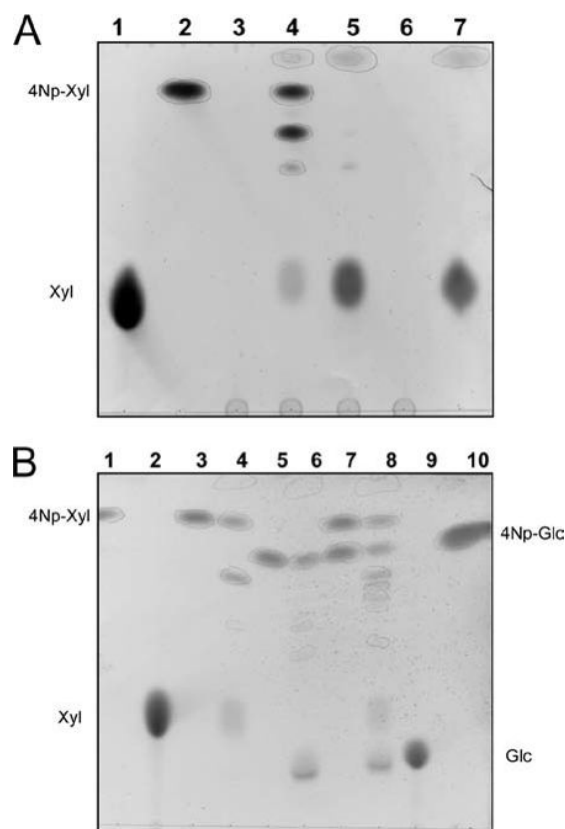


FIGURE 3. **Reaction products of SSO1353.** Thin layer chromatography of the transxylosylation reactions with 4Np-Xyl. (A) In each lane were loaded 20  $\mu$ l of the following reaction mixtures: lane 1, xylose standard 25 mM; lane 2, blank mixture containing 50 mM sodium citrate buffer pH 5.5, 4Np-Xyl 5 mM; lane 3, SSO1353 after TP85  $^{\circ}$ C (30  $\mu$ g); lane 4, reaction mixture same as lane 2 with added SSO1353 after TP85  $^{\circ}$ C (150  $\mu$ g); lane 5, sample loaded in lane 4 with added S $\beta$ -gly (50  $\mu$ g); lane 6, S $\beta$ -gly (25  $\mu$ g); lane 7, sample loaded in lane 2 with added S $\beta$ -gly (25  $\mu$ g). Transxylosylation reactions with 4Np-Xyl and 4Np-Glc. (B) In each lane were loaded 20  $\mu$ l of the following reaction mixtures: lane 1, 4Np-Xyl 5 mM; lane 2, xylose standard 25 mM; lane 3, blank mixture containing 50 mM sodium citrate buffer, pH 5.5, 4Np-Xyl 5 mM; lane 4, reaction mixture same as lane 3 with added pure SSO1353 (7  $\mu$ g); lane 5, blank mixture containing 50 mM sodium citrate buffer, pH 5.5, 4Np-Glc 5 mM; lane 6, same as lane 5 with added pure SSO1353 (7  $\mu$ g); lane 7, blank mixture containing 50 mM sodium citrate buffer, pH 5.5, 4Np-Xyl 5 mM and 4Np-Glc 5 mM; lane 8, same as lane 7 with added pure SSO1353 (7  $\mu$ g); lane 9, glucose standard 25 mM; lane 10, 4Np-Glc 5 mM.

ate (glycosylation step) (Fig. S2). In the second step (de-glycosylation step), the group previously acting as an acid now works as a base catalyst deprotonating the water and resolving the glycosyl enzyme intermediate. Both steps proceed via transition states with substantial oxocarbenium ion character (24).

The identification of key active site residues in a glycoside hydrolase is crucial to determine the catalytic machinery for the classification of this class of enzymes (5, 25, 26). These residues can be identified by using several different techniques, site-directed mutagenesis followed by kinetic analysis of the mutants being one of the approaches most used. Briefly, con-

FIGURE 1. **Phylogenetic tree of family GH116 using Ward hierarchical clustering distances.** The leaves of the tree indicate the organism genus and species information (e.g. *Homo sapiens* corresponds to *Homo sapiens*), the gene or Locus name, the EC activities if characterized experimentally, and a database accession number. Tree branches were colored according to identified significant subgroups using Dendroscope 2.3 (21).

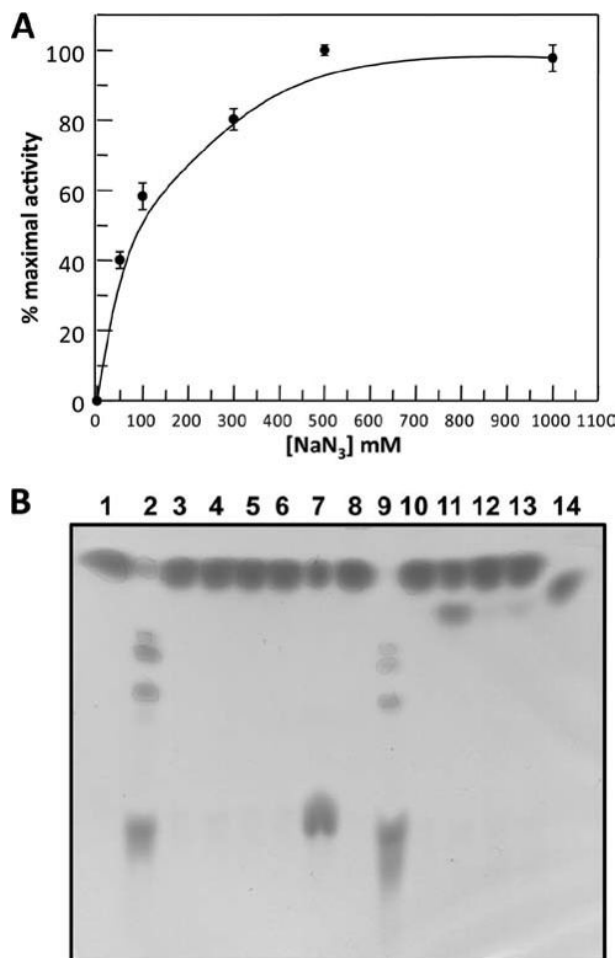
## Characterization of an Archaeal $\beta$ -Glycosidase

served aspartic/glutamic acid residues identified by sequence analysis are mutated with non-nucleophilic amino acids; the reduction or even abolition of the enzymatic activity is a strong indication that the mutation removed catalytic residues. The activity of the mutants can be chemically rescued in the presence of external nucleophiles such as sodium azide. The characterization of the anomeric configuration of the glycosyl-azide products allows the assignment of the mutated residue as the nucleophile or the acid/base of the reaction (Fig. S2B) (27).

We aligned the amino acid sequence of SSO1353 to eight other hypothetical proteins identified by BLAST analysis with an identity  $\geq 22\%$ . The multi-alignment led to the identification of 15 Asp/Glu residues highly conserved (Fig. S3); among these, Glu-335, Asp-406, and Asp-426 were invariant and, together with Asp-458, were mutated by site-directed mutagenesis obtaining the mutants E335G, D406G, D458G, and D462G. These SSO1353 mutants were expressed and purified as described above. During this procedure the proteins showed identical behavior, suggesting that the mutations did not affect the stability of the enzymes. These purification steps yielded proteins with similar concentrations and purification degrees (Fig. S4). The mutants assayed at 65 °C on 2Np-Glc 40 mM in 50 mM sodium citrate buffer pH 5.5 were completely inactive indicating that the mutations affected the catalytic machinery of SSO1353. When 1 M sodium azide was included in the assay on 40 mM 2Np-Glc we observed the reactivation of the D462G mutant, which showed a specific activity of 0.5 units  $\text{mg}^{-1}$ , which is about 7-fold lower than that of the wild type (3.6 units  $\text{mg}^{-1}$ ) assayed in the same conditions. Instead, the external ion did not modify the specific activity of the wild type and did not reactivate the E335G, D406G, and D458G mutants.

The mutant D462G was assayed at standard conditions on 40 mM 2Np-Glc in the presence of increasing concentrations of sodium azide: the maximal activity was observed at 0.5 M sodium azide (Fig. 4A). At these conditions the kinetic constants were  $k_{\text{cat}}$  of  $0.64 \pm 0.1 \text{ s}^{-1}$ ,  $K_m$  of  $16.2 \pm 6 \text{ mM}$ , and  $k_{\text{cat}}/K_m$  of  $0.04 \text{ s}^{-1} \text{ mM}^{-1}$ , showing that D462G maintained a similar affinity for the substrate, but a specificity constant 10-fold lower than the wild type. Reaction mixtures prepared at these conditions and containing the wild type and the mutants were analyzed by TLC after prolonged incubation. The D462G produced a novel compound, which was observed only in trace amounts with the other mutants. Instead, the wild type completely converted the substrate producing transglycosylation products (Fig. 4B). D462G reaction mixtures in preparative scale allowed the isolation and structural characterization of this product that was unequivocally identified as  $\beta$ -glucosyl azide. The reactivation in the presence of the external ion and the anomeric configuration of this product strongly indicate that Asp-462 is the acid/base of the reaction.

**Identification of the Catalytic Nucleophile of SSO1353**—To identify the nucleophile of the reaction we used the mechanism-based inhibition approach in combination with nano-electrospray ionization tandem mass spectrometry (nano-ESI-MS/MS) analysis. Mechanism-based inhibitors are ligands that bind to the active site by competing with the substrate of retaining glycosidases and require mechanism-based activation to react covalently with the enzyme (for a review see Ref. 28). One



**FIGURE 4. Chemical rescue of the activity of SSO1353 mutants.** (A) Dependence of activity of D462G mutant on different concentrations of sodium azide. (B) TLC analysis of the reaction mixtures of SSO1353 wild type and mutants in the absence (lanes 1–6) and in the presence of 0.1 M sodium azide (lanes 8–13). Standard assays were performed overnight at 65 °C on 40 mM 2Np-Glc by using 11  $\mu\text{g}$  of enzyme. Lane 1, blank with no enzyme; lane 2, wild type; lane 3, E335G mutant; lane 4, D462G; lane 5, D458G; lane 6, D406G; lane 7, standards (2Np-Glc and Glc); lane 8, blank with no enzyme; lane 9, wild type; lane 10, E335G; lane 11, D462G; lane 12, D458G; lane 13, D406G; lane 14,  $\beta$ -glucosyl-azide standard.

group of these inhibitors includes activated 2-deoxy-2-fluoro-glycosides; the presence of fluorine substituent at C2 slows both the glycosylation and the deglycosylation steps of the reaction by destabilizing the transition states. The incorporation of good leaving groups (as 2,4-dinitrophenol or fluoride) accelerates the glycosylation step relative to the deglycosylation step of the reaction with the effect that the incubation of the enzyme with its corresponding 2-deoxy-2-fluoro-glycosides results in a time dependent inactivation with the accumulation of the 2-deoxy-2-fluoro-glycoside enzyme intermediate. Consequently, the nucleophile of the reaction labeled with the inhibitor can be identified by mass spectrometry (29).

Time-dependent inactivation of SSO1353 was observed upon incubation of the enzyme with 2,4DNp-2F-Glc (Fig. 5). Inhibition was incomplete after 4 h of incubation (about 40%)



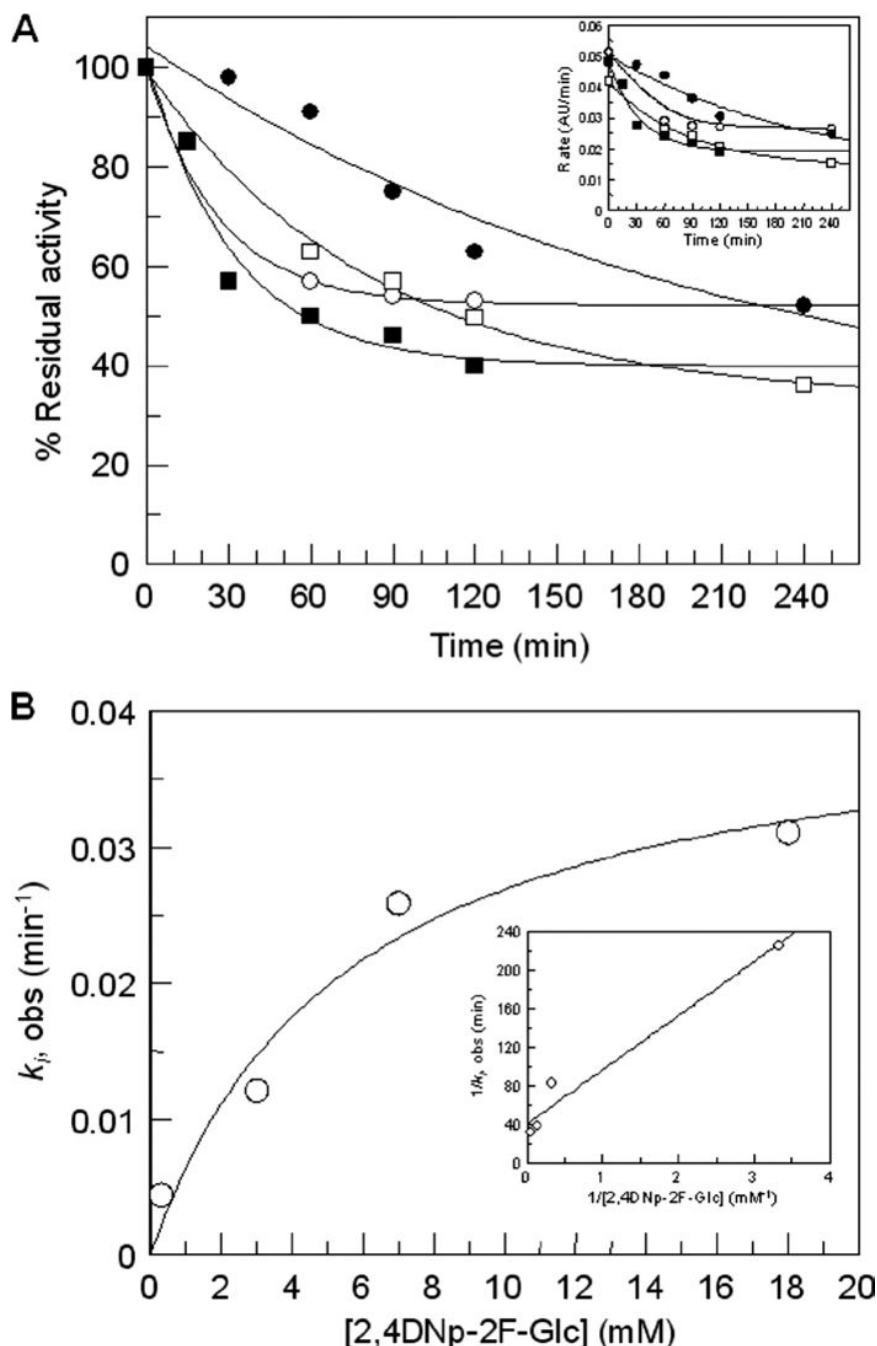


FIGURE 5. Time-dependent inactivation of SSO1353 using 2,4DNp-2F-Glc inhibitor. A plot of % residual activity versus time at four 2,4DNp-2F-Glc concentrations (● 0.3 mM, □ 3 mM, ○ 7 mM, and ■ 18 mM) is reported in A with the plot of rate versus time shown as inset.  $k_{\text{obs}}$  were obtained from fitting the curves in A to a single exponential decay with offset because time-dependent inactivation did not decay to zero, and plotted versus inhibitor concentration (B), to determine  $K_i$  and  $k_i$ . The reciprocal plot is shown as inset in B.

even at the highest concentration of inhibitor used (18 mM). This is not surprising as the catalytic competence of GH inactivated by mechanism-based inhibitors, occurring via turnover of the intermediate via hydrolysis or transglycosylation, has been well documented (28). At these conditions we obtained

the following inactivation parameters:  $k_i = (6.9 \pm 1.3) \times 10^{-4} \text{ s}^{-1}$ ;  $K_i = 5.5 \pm 2.7 \text{ mM}$ ;  $k_i/K_i = 1.2 \times 10^{-4} \text{ s}^{-1} \text{ mM}^{-1}$ .

SSO1353 samples incubated in the absence and the presence of 2.9 mM 2,4DNp-2F-Glc for 2 h were analyzed by single-stage nano-ESI-MS to monitor alteration in the molecular mass of the protein. Nano-ESI mass spectra of intact proteins yield series of multiply charged molecular ion peaks with 40–100 positive charges under the experimental condition applied. Molecular mass of SSO1353 in the absence of inhibitor was measured to be  $75914 \pm 6 \text{ Da}$ , which is comparable to the theoretical average molecular mass (75907.7 Da) within experimental error (0.009%) (Fig. S5A). After inhibition, molecular ion peaks shift toward higher  $m/z$  values leading a molecular mass of  $76077 \pm 5 \text{ Da}$  and accounting for  $163 \pm 5.5 \text{ Da}$  difference between the two species (Fig. S5B). To gain further evidence and a more detailed structural insight into the site-directed inhibition, samples were proteolytically digested by pepsin and the resulting peptide mixtures were analyzed by nano-HPLC-ESI-MS/MS in IDA mode. Based on MS/MS sequence data, 91 and 87% protein sequence coverage were respectively obtained in the absence and in the presence of inhibitor (Table S1A and Table S1B respectively). Interestingly, in the inhibited sample five peptides comprising residues 332–345, 332–343, 332–347, 332–348, and 332–349 showed considerable decrease in intensity, and in the same time, six new peptide molecular ions appeared in the corresponding scans (Table 2). The peptide molecular ion pairs corresponding to the normal and the modified sequences eluted at the same retention time and thus they were detected in the same survey scan in the sample containing 2,4DNp-2F-Glc. Therefore, they are likely due to in-source ion fragmentation process indicating a relatively labile bond between amino acid and inhibitor. These peptides showed an increase of 164.05 Da in molecular mass which corresponds well to the difference between unmodified and 2F-Glc



## Characterization of an Archaeal $\beta$ -Glycosidase

**TABLE 2**

**Peptide mapping of SSO1353**

Characteristic peptide molecular ions containing residue E at position 335 observed during nano-HPLC-ESI-MS/MS IDA analyses of SSO1353 incubated in the absence and in the presence of 2,4Dnp-2F-Glc inhibitor and digested by pepsin. Peptide sequences (both unmodified and modified) were elucidated by the interpretation of nano-ESI-MS/MS spectra acquired on the doubly charged ( $z = 2$ ) precursor ions (Fig. 4). Modification corresponds to the covalent attachment of 2F-Glc ligand at E335 (indicated in the sequence as E<sup>\*</sup>).

From-To	Unmodified and modified peptide sequences	Mw (calc.)	SSO1353			SSO1353 inhibited		
			Rt	m/z	Intensity	Rt	m/z	Intensity
			<i>min</i>		<i>cps</i>	<i>min</i>		<i>cps</i>
332–345	AIYEAPQNCPLYGT	1538.708	20.7	770.36	26	21.2	770.36	10
	AIYE <sup>*</sup> APQNCPLYGT	1703.188	–	n.d. <sup>a</sup>	–	–	852.39	34
332–343	AIYEAPQNCPLY	1380.638	21.8	691.33	30	22.0	691.33	17
	AIYE <sup>*</sup> APQNCPLY	1544.689	–	n.d.	–	–	773.35	44
332–347	AIYEAPQNCPLYGTIG	1708.813	23.2	855.42	42	23.2	855.42	16
	AIYE <sup>*</sup> APQNCPLYGTIG	1873.293	–	n.d.	–	–	937.44	37
332–348	AIYEAPQNCPLYGTIGA	1779.85	23.5	890.94	280	23.7	890.93	63
	AIYE <sup>*</sup> APQNCPLYGTIGA	1943.898	–	n.d.	–	–	972.96	240
332–349	AIYEAPQNCPLYGTIGAC	1882.859	24.8	942.46	320	24.6	942.44	106
	AIYE <sup>*</sup> APQNCPLYGTIGAC	2046.07	–	n.d.	–	–	1024.46	384
332–340	AIYEAPQNC	1007.447	–	n.d.	–	25.2	n.d.	–
	AIYE <sup>*</sup> APQNC	1171.495	–	n.d.	–	–	586.76	94

<sup>a</sup> n.d., not determined.

modified peptides, and indicated that the ligand was likely bound to one of the amino acids present in peptide 332–349. To confirm the site of modification, nano-ESI-MS/MS spectra of the unmodified/modified peptide pairs were analyzed (Table 2, Fig. 6). MS/MS spectra show a very similar fragmentation pattern yielding characteristic b-type N-terminal fragment ions at the low  $m/z$  range. Based on these ions, and in particular, on the appearance of  $b_n^*$  ( $n \geq 3$ ) modified fragment ions at  $m/z$  641.28 ( $b_3^*$ ), 712.32 ( $b_4^*$ ) and 809.37 ( $b_5^*$ ) in the inhibited sample, modification was unequivocally localized on amino acid Glu-335. Therefore, it was concluded that Glu-335 is the nucleophile of the reaction of SSO1353. Though we had no direct evidence from the chemical rescue experiment, we deduce that the invariant residue Asp-462 is the acid/base of the reaction.

## DISCUSSION

We report here the molecular cloning, the expression in *E. coli* and the functional characterization of the product of the gene *ssol353* from the hyperthermophilic archaeon *S. solfataricus*. The molecular characterization revealed the specificity of the enzyme for gluco- and -xylosides  $\beta$ -bound to hydrophobic groups that are hydrolyzed by following a retaining reaction mechanism. In addition, site-directed mutagenesis of conserved glutamic/aspartic amino acids and the chemical rescue of the  $\beta$ -glycosidase activity of the mutants, combined with the use of mechanism based inhibitors and mass spectrometric analysis, allowed us to identify Asp-462 and Glu-335 as the acid/base and the nucleophile of the reaction, respectively. Mutagenic studies also suggested that Asp-406 and Asp-458 residues play a role in catalysis, but elucidation of their function requires further investigations. Amino acid sequence analysis showed that SSO1353 shared identity with other hypothetical proteins and, remarkably, with eukaryotic non-lysosomal bile acid  $\beta$ -glucosidases.

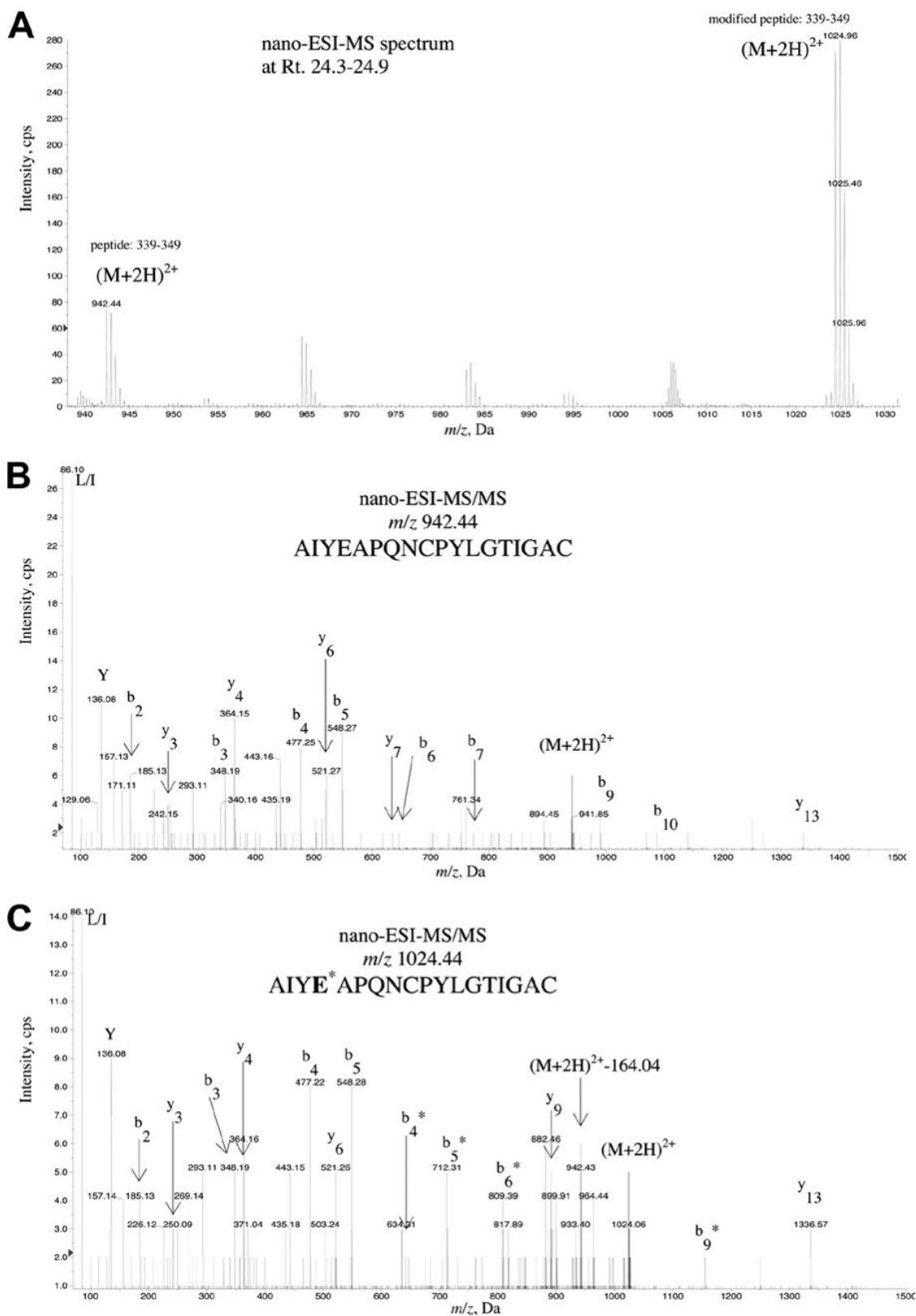
So far, SSO1353 was not assigned to a defined glycoside hydrolase family in the carbohydrate active enzyme database. On the basis of our findings we propose that SSO1353 and its homologs define a new sequence-based family, namely GH116, which presently includes enzymes with  $\beta$ -glucosidases (EC 3.2.1.21),  $\beta$ -xylosidases (EC 3.2.1.37), or glucocerebrosidases

(EC 3.2.1.45) activity. As for the other GH families, the retaining reaction mechanism and the catalytic role for the acid/base and the nucleophile, experimentally determined here, can be easily extended to all the enzymes belonging to this new family.

Interestingly, all the archaeal putative enzymes belonging to this new family are from Crenarchaea, and the vast majority originates from the genus *Sulfolobus*. A PSI-BLAST search conducted using SSO1353 as the query sequence retrieved (with low scores) uncharacterized bacterial glycosidases belonging to families GH15, GH63, and GH78. The latter families include mainly glucoamylases,  $\alpha$ -glucosidases, and  $\alpha$ -L-rhamnosidases, respectively, and are characterized by an  $(\alpha/\alpha)_6$  fold. Although SSO1353 is inactive on  $\alpha$ -glycosides, this perhaps hints at structural similarities with enzymes from family GH116. Similar structural similarity between  $(\alpha/\alpha)_6$  fold glycoside hydrolase families degrading both  $\alpha$  and  $\beta$  glycosidic bonds have already been described (30).

The phylogenetic analysis (Fig. 1) shows that sequences from the new family GH116 can be subdivided into two major groups, one containing sequences from Archaea and another one composed mostly of sequences from Cyanobacteria and Eukaryotes. The archaeal subgroup can be further subdivided into at least two subgroups, in which, interestingly, all the archaeal homologs of SSO1353 are present as multiple copies in the genomes of *Caldivirga maquilingensis*, *S. tokodaii*, *S. solfataricus*, and in the six strains of *S. islandicus*. The *ssol353* homologs with identity  $>80\%$ , lie downstream of genes encoding endoglucanases, and, interestingly, in *S. solfataricus*, this gene arrangement occurs twice. Presumably, the  $\beta$ -glycosidase activity of SSO1353 is involved, in combination with the secreted endoglucanase, in the degradation of exogenous glucans used as carbon energy source or, possibly, of the exopolysaccharides (EPS) that are produced by *S. solfataricus* itself (31–33). Other *ssol353* homologs, with identities in the range 21–33% exemplified by *ssol2674* and *ssol3039* in *S. solfataricus*, flank a putative peptidase or a putative gluconolactonase, respectively. These two other subgroups of enzymes similar to SSO1353 are present also in *C. maquilingensis*, *S. tokodaii*, and *S. islandicus* showing a remarkable identity ( $>80\%$ ) within each

# Characterization of an Archaeal $\beta$ -Glycosidase



## Characterization of an Archaeal $\beta$ -Glycosidase

Pan	-----LPEELG-RNMCH-----LRPTLRDYGRFGYLBQ	456
Homo	-----LPEELG-RNMCH-----LRPTLRDYGRFGYLBQ	529
Ciona	-----NGKVPETLQNTTKVD-----SCDFLKEYGKFAYLBQ	498
Arabidopsis	GLKNDIDVPHQNDTAVSVLEKMASTLEELHASTTSNSAFGKLLLEEENIGHFYLBI	535
Sulfolobus	-----EKGGRFAIYBAP	337
	: *:* *	
Pan	EYRMYN--TYDVHFYASFALIMLWPKLELSLQYDMALATLREDLTRRRYLMSGVMAVVKR	514
Homo	EYRMYN--TYDVHFYASFALIMLWPKLELSLQYDMALATLREDLTRRRYLMSGVMAVVKR	587
Ciona	EYKMYN--TYDVHFYASIALAYLWPKLELSVQYDIATISILHNSNPQPHKYLMDGVTA PVKT	556
Arabidopsis	EYRMWN--TYDVHFYASFALVMLFPKLELSIQDFAAAVMLHDP TKVKTLSEGGQVVRKV	593
Sulfolobus	QNCPLYGTIGACYEFGSLPVILMFPELEKSLKLLIRHIRE-----	379
	: : : : *:* * :	
Pan	RNVIPHDIGDPDDEPWLVRNAYLIHDTADWKDLNLKFLVQVYRDYLLTGDNFLKDMWPV	574
Homo	RNVIPHDIGDPDDEPWLVRNAYLIHDTADWKDLNLKFLVQVYRDYLLTGDNFLKDMWPV	647
Ciona	PNVVPDVGCPDEDEPWLVRNTYFVHDTADWRDLNPKFVLQAYRDYITKIDIFLKAMWPI	616
Arabidopsis	LGAVPHDLG--INDPWFEVNGYTLHNTDRWKDLNPKFVLQVYRDVVATGDKKFASAVWPS	651
Sulfolobus	-GYVPHDLGYHSLDSPIDG---TSPPRWKDMNPSLILLVYRYFKFTNDIEFLKEVYPI	434
	. : *:* * : : : *:* * : * * . * . : *	
Pan	CLAVMESEMKFDKDHGLIENGGYADQTYDGVWTTGPSAYCGGLWLA AVAVMVQMAALCG	634
Homo	CLAVMESEMKFDKDHGLIENGGYADQTYDGVWTTGPSAYCGGLWLA AVAVMVQMAALCG	707
Ciona	CKIVMEQSMRHKDNDGLIENGGAADQTFDGCWVTGPSAYCGGLWLAALRCMEEAADILH	676
Arabidopsis	VYVAMAYMAQFDKDGDMIEGFPDQTYDTWSAGVSAYCGGLWVAALQAASALARVVG	711
Sulfolobus	LVKVMDELRLQCKGNLPFMEG--EMDNALFATIIKGHDSYTSLSLIGSLIAMREIAKLVG	492
	. * : * . : : * : * : * : * : * : *	

FIGURE 7. Multi-alignment of SSO1353 with glucosylceramidases. Invariant residues are indicated with "###"; increased level of conservation is indicated with "\*" and "." The residues corresponding to the nucleophile Glu-335 and acid/base Asp-462 of SSO1353 are boxed. Pan is XP\_001167952.1 from *Pan troglodytes*; Homo is NP\_065995.1 from *Homo sapiens*; Ciona is XP\_002127036.1 from *Ciona intestinalis*; Sulfolobus is SSO1353 from *S. solfataricus* P2.

subgroup. The observation that the archaeal  $\beta$ -glycosidases from this novel GH family can be subgrouped according to their identity suggests they are present in multiple copies for functional purposes, possibly, for the degradation/modification of different substrates. A more detailed characterization of these enzymes is needed to understand their function *in vivo*.

The other major subdivision of the family is prone to be subdivided into several subgroups, one containing sequences from Cyanobacteria, the other having plant, animal, and mixed bacterial subdivisions. One of the members of the animal subgroup in this newly proposed family is human non-lysosomal glucosylceramidase or  $\beta$ -glucosidase 2 (GBA2). This enzyme, previously described as bile acid  $\beta$ -glucosidase (34), is involved in the catabolism of glucosylceramide, which is then converted to sphingomyelin (6). Glucocerebrosidases are important enzymes involved in the metabolism of gangliosides and globosides. Deficiency of this enzymatic activity is the cause of the most common lysosomal storage disorder named Gaucher disease (35) resulting from a defect in the lysosomal acid  $\beta$ -glucosidase (GBA1) belonging to GH30. This deficiency leads to the accumulation of glycosylceramides in certain organs, typically spleen, kidney, lungs, brain, and bone marrow (36). The finding that other cell types of Gaucher patients did not show accumulation of glycosylceramides suggested the existence of an alternative catabolic pathway that later was demonstrated to be catalyzed by GBA2 (6). This enzyme is ubiquitously expressed and it is associated to the cell surface. GBA2 is inactive on MU-Xyl,

is inhibited by hydrophobic deoxynojirimycin (DNJ), and it is relatively insensitive to CBE (6, 34, 37). In humans, no known pathologies related to defects of GBA2 have been reported so far while only in certain mice strains treatments with NB-DNJ or *gba2* gene knock-outs led to impaired spermatogenesis (38). However, such deleterious effects were not observed in other organisms including humans (6, 39, 40). These studies demonstrate the importance of understanding at the molecular level the reaction mechanism and the catalytic machinery of carbohydrate active enzymes for the development of specific inhibitors for bio-medical applications. The experimental identification of the catalytic amino acids of SSO1353 reported here, allows to easily identifying the catalytic machinery of human GBA2 despite the low sequence identity (18%) between the two enzymes. In GBA2 the nucleophile and the acid/base of the

reaction are Glu-528 and Asp-678, respectively, which, as observed in a multi-alignment of putative glucocerebrosidases from mammals, plants, and tunicates belonging to this new GH family, are located in two conserved motifs (Fig. 7). In particular, amino acids with hydrophobic side chains are almost invariant in the position preceding the catalytic glutamic and aspartic acids in the enzymes belonging to the new family GH116 (Fig. S3 and Fig. 7). Our findings can now allow the planning of more detailed site-directed mutagenesis studies to better understand the molecular bases of the substrate recognition of GBA2.

SSO1353 has substrate specificity and inhibitor sensitivity slightly different from those of GBA2. In fact, the archaeal enzyme can hydrolyze both aryl  $\beta$ -gluco and  $\beta$ -xylosides and it is inhibited with mM affinity by both NB-DNJ and CBE. Instead, GBA2 is inactive on MU-Xyl and it is relatively insensitive to CBE (6). These differences presumably reflect the different function of the two enzymes *in vivo*: the wider substrate specificity of the archaeal enzyme might allow to degrade a variety of substrates ensuring an efficient availability of sugars as energy source while GBA2 is involved in a well defined catabolic pathway. The purification of GBA2 is made difficult by its instability to detergents precluding its production in abundant and homogeneous form (6). Instead, robust GBA2 homologs from hyperthermophilic Archaea can be more easily expressed and purified from conventional

FIGURE 6. Identification of the catalytic nucleophile site of SSO1353 by nano-HPLC-ESI mass spectrometry. (A) Nano-ESI mass spectrum of peptides eluted at 24.3–24.9 min of SSO1353 incubated with 2,4DNp-2F-Glc for 2 h. Doubly charged peptide ions at  $m/z$  942.44 and 1024.46 correspond to the unmodified and the modified peptide 339–349, respectively. Tandem mass spectra on the unmodified (B) and modified peptides 339–349 (C) reveal modification on E4 residue.

hosts allowing more simple structural studies that might be easily extended to the human counterpart.

*Acknowledgments*—We thank Gerard W. Dougherty for the editing of the manuscript and Giovanni D'Angelo and Antonella De Matteis for the generous gift of the gangliosides. The IBP-CNR belongs to the Centro Regionale di Competenza in Applicazioni Tecnologico-Industriali di Biomolecole e Biosistemi.

## REFERENCES

- Varki, A. (1993) *Glycobiology* **3**, 97–130
- Caramelo, J. J., and Parodi, A. J. (2007) *Semin Cell Dev. Biol.* **18**, 732–742
- Saxon, E., and Bertozzi, C. R. (2001) *Annu. Rev. Cell Dev. Biol.* **17**, 1–23
- DeMarco, M. L., and Woods, R. J. (2008) *Glycobiology* **18**, 426–440
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) *Nucleic Acids Res.* **37**, D233–8
- Boot, R. G., Verhoeck, M., Donker-Koopman, W., Strijland, A., van Marle, J., Overkleeft, H. S., Wennekes, T., and Aerts, J. M. (2007) *J. Biol. Chem.* **282**, 1305–1312
- van Weely, S., Brandsma, M., Strijland, A., Tager, J. M., and Aerts, J. M. (1993) *Biochim. Biophys. Acta* **1181**, 55–62
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
- Mackenzie, L. F., Brooke, G. S., Cutfield, J. F., Sullivan, P. A., and Withers, S. G. (1997) *J. Biol. Chem.* **272**, 3161–3167
- Leatherbarrow, R. J. (1992) *Erithacus Software Ltd.*, Staines, UK
- Shaikh, F. A., Müllegger, J., He, S., and Withers, S. G. (2007) *FEBS Lett.* **581**, 2441–2446
- Perugino, G., Falcicchio, P., Corsaro, M. M., Matsui, I., Parrilli, M., Rossi, M., and Moracci, M. (2006) *Biotrans.* **24**, 23–29
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
- Edgar, R. C. (2004) *Nucleic Acids Res.* **32**, 1792–1797
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) *Bioinformatics* **25**, 1189–1191
- Le, S. Q., and Gascuel, O. (2008) *Mol. Biol. Evol.* **25**, 1307–1320
- Ward, J. H. (1963) *J. Am. Stat. Assoc.* **58**, 236–244
- Maurelli, L., Giovane, A., Esposito, A., Moracci, M., Fiume, I., Rossi, M., and Morana, A. (2008) *Extremophiles*. **12**, 689–700
- Torarinsson, E., Klenk, H. P., and Garrett, R. A. (2005) *Environ Microbiol.* **7**, 47–54
- Huang, Y., Krauss, G., Cottaz, S., Driguez, H., and Lipps, G. (2005) *Biochem. J.* **385**, 581–588
- Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., and Rupp, R. (2007) *BMC Bioinformatics* **8**, 460
- Moracci, M., Ciaramella, M., and Rossi, M. (2001) *Methods Enzymol.* **330**, 201–215
- Bock, K., and Pedersen, C. (1983) *Adv. Carbohydr. Chem. Biochem.* **41**, 27–66
- Koshland, D. E. (1953) *Biol. Rev. Camb. Philos. Soc.* **28**, 416–436
- McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol.* **4**, 885–8892
- Zeche, D. L., and Withers, S. G. (2000) *Acc. Chem. Res.* **33**, 11–18
- Ly, H. D., and Withers, S. G. (1999) *Annu. Rev. Biochem.* **68**, 487–522
- Withers, S. G., and Aebersold, R. (1995) *Protein Sci.* **4**, 361–372
- Williams, S. J., and Withers, S. G. (2000) *Carbohydr. Res.* **327**, 27–46
- Stam, M. R., Blanc, E., Coutinho, P. M., and Henrissat, B. (2005) *Carbohydr. Res.* **340**, 2728–2734
- Elferink, M. G., Albers, S. V., Konings, W. N., and Driessen, A. J. (2001) *Mol. Microbiol.* **39**, 1494–1503
- Nicolaus, B., Manca, M., Romano, I., and Lama, L. (2003) *FEMS Microbiol. Lett.* **109**, 203–206
- Zolghadr, B., Klingl, A., Koerd, A., Driessen, A. J., Rachel, R., and Albers, S. V. (2010) *J. Bacteriol.* **192**, 104–110
- Matern, H., Boermans, H., Lottspeich, F., and Matern, S. (2001) *J. Biol. Chem.* **276**, 37929–37933
- Grabowski, G. A. (2008) *Lancet* **372**, 1263–1271
- Grabowski, G. A., Gatt, S., and Horowitz, M. (1990) *Crit. Rev. Biochem. Mol. Biol.* **25**, 385–414
- Overkleeft, H. S., Renkema, G. H., Neele, J., Vianello, P., Hung, I. O., Strijland, A., van der Burg, A. M., Koomen, G. J., Pandit, U. K., and Aerts, J. M. (1998) *J. Biol. Chem.* **273**, 26522–26527
- van der Spoel, A. C., Jeyakumar, M., Butters, T. D., Charlton, H. M., Moore, H. D., Dwek, R. A., and Platt, F. M. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17173–17178
- Amory, J. K., Muller, C. H., Page, S. T., Leifke, E., Pagel, E. R., Bhandari, A., Subramanyam, B., Bone, W., Radlmaier, A., and Bremner, W. J. (2007) *Hum. Reprod.* **22**, 702–707
- Bone, W., Walden, C. M., Fritsch, M., Voigtmann, U., Leifke, E., Gottwald, U., Boomkamp, S., Platt, F. M., and van der Spoel, A. C. (2007) *Reprod. Biol. Endocrinol.* **5**, 1–13

## A novel $\alpha$ -D-galactosynthase from *Thermotoga maritima* converts $\beta$ -D-galactopyranosyl azide to $\alpha$ -galacto-oligosaccharides

Beatrice Cobucci-Ponzano<sup>1</sup>, Carmela Zorzetti<sup>1</sup>, Andrea Strazzulli<sup>1</sup>, Sara Carillo<sup>2</sup>,  
Emiliano Bedini<sup>2</sup>, Maria Michela Corsaro<sup>2</sup>, Donald A. Comfort<sup>3,4</sup>, Robert M. Kelly<sup>3</sup>,  
Mosè Rossi<sup>1,5</sup>, and Marco Moracci<sup>1,\*</sup>

<sup>1</sup>Institute of Protein Biochemistry – Consiglio Nazionale delle Ricerche, Via P. Castellino 111, 80131, Naples, Italy. <sup>2</sup>Dipartimento di Chimica Organica e Biochimica, “Università di Napoli Federico II”, Complesso Universitario di Monte S. Angelo, via Cinthia 4, 80126 Naples, Italy. <sup>3</sup>Department of Chemical and Biomolecular Engineering, North Carolina State University, EB-1, 911 Partners Way, Raleigh, NC 27695-7905, USA. <sup>4</sup>Department of Chemical and Materials Engineering, University of Dayton, 300 College Park, Dayton OH 45469-0246, USA (current address) <sup>5</sup>Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli “Federico II”, Complesso Universitario di Monte S. Angelo, Via Cinthia 4, 80126 Naples, Italy.

\*Author to whom correspondence should be addressed: Dr. Marco Moracci. Tel +390816132271; Fax +390816132277; e-mail: m.moracci@ibp.cnr.it

This submission includes supplementary Table S1 and Figure S1.

## Abstract

The large-scale production of oligosaccharides is a daunting task, hampering the study of the role of glycans *in vivo* and the testing of the efficacy of novel glycan-based drugs.

Glycosynthases, mutated glycosidases that synthesize oligosaccharides in high yields, are becoming important chemo-enzymatic tools for the production of oligosaccharides. However, while  $\beta$ -glycosynthase can be produced with a rather well established technology, examples of  $\alpha$ -glycosynthases are thus far limited only to enzymes from families GH29, GH31 and GH95.  $\alpha$ -L-Fucosynthases from GH29 use convenient glycosyl azide derivatives as a strategic alternative to glycosyl fluoride donors. However, the general applicability of this method to other  $\alpha$ -glycosynthases is not trivial and remains to be confirmed. Here,  $\beta$ -D-galactopyranosyl azide was converted to  $\alpha$ -galacto-oligosaccharides with good yields and high regioselectivity, catalyzed by a novel  $\alpha$ -galactosynthase based on the GH36  $\alpha$ -galactosidase from the hyperthermophilic bacterium *Thermotoga maritima*. These results open a new avenue to the practical synthesis of biologically interesting  $\alpha$ -galacto-oligosaccharides and demonstrate more widespread use of  $\beta$ -glycosyl-azide as donors confirming their utility to expand the repertoire of glycosynthases.

**Keywords:**  $\alpha$ -Galacto-oligosaccharides/glycosynthase/ $\alpha$ -gal epitope/carbohydrate synthesis/protein engineering



## Introduction

Reliable methods for large-scale production of oligosaccharides are currently unavailable despite the considerable interest in these molecules. Among other restrictions, this hampers development of technical strategies needed to prove their role *in vivo* and the efficacy of glycan-based drugs (Brown *et al.* 2008; Drake *et al.* 2010; Kiessling and Splain 2010). Oligosaccharide synthesis is challenging because of the diverse stereochemistry of the monosaccharide building blocks, the enormous number of intersugar linkages that they can form, and the absence, in contrast to nucleic acids and protein synthesis, of biological processes ruled by universal conserved codes (Gabijs *et al.* 2004). This challenge is currently tackled with a variety of approaches, often employed concurrently, and aimed at automated carbohydrate synthesis (Seeberger 2008). Classical chemical synthesis, although powerful, is hampered by laborious regio- and stereochemical control, while the practical use of highly specific and efficient glycosyltransferases (GT) is delayed by the costs of substrates and catalysts (Hancock *et al.* 2006). Glycoside hydrolases (GH), working in transglycosylation mode, are an interesting alternative (McCarter and Withers 1994). These enzymes, which retain the anomeric configuration of the substrate in the product, follow a double displacement mechanism, in which a glycosyl-intermediate is formed and subsequently hydrolyzed (Figure 1A) (Koshland 1953). In the presence of high concentrations of acceptors different from water, transglycosylation occurs. The main drawback of this approach is that the newly formed product is a substrate of the enzyme and can be hydrolyzed with reduced final yields. To overcome these limitations, glycosynthases, a new class of mutant glycosidases in which the active site nucleophile is replaced with a non-nucleophilic residue, were introduced (Mackenzie *et al.* 1998; Moracci *et al.* 1998; Malet and Planas 1998). These enzymes catalyze the synthesis of oligosaccharides starting from substrates with inverted anomeric configuration when compared to the original substrate (Figure 1B, *inverting* glycosynthases)

(Mackenzie *et al.* 1998; Malet and Planas 1998), or by exploiting small external nucleophiles (Figure 1B, *retaining* glycosynthases) (Moracci *et al.* 1998). Several glycosidases from a variety of families of the Carbohydrate Active enZYme classifications (CAZy) <http://www.cazy.org/> (Cantarel *et al.* 2009) (namely, GH1, GH2, GH5, GH7, GH8, GH10, GH16, GH17, GH26, GH52, and GH85 (Perugino *et al.* 2004; Honda and Kitaoka 2006; Shaikh and Withers 2008), hydrolyzing  $\beta$ -O-linked sugars (equatorial) have been converted into efficient glycosynthases. In contrast, examples of  $\alpha$ -glycosynthases have thus far been limited to  $\alpha$ -L-fucosidases from families GH29 and GH95 (Cobucci-Ponzano *et al.* 2009; Wada *et al.* 2008) and to an  $\alpha$ -glucosidase from family GH31 (Okuyama *et al.* 2002), thereby limiting the access of the glycosynthase approach to  $\alpha$ -fucosylated,  $\alpha$ -galactosylated, and  $\alpha$ -mannosylated oligosaccharides, which are implicated in important biological phenomena (Vanhooren and Vandamme 1999; Macher and Galili 2008). Along these lines, we recently reported that two phylogenetically unrelated  $\alpha$ -L-fucosidases could be converted in efficient  $\alpha$ -fucosynthases by reactivating the nucleophile mutants with  $\beta$ -fucosyl-azide (Figure 1C), as an alternative to the classical  $\beta$ -fluoride derivative donor (Cobucci-Ponzano *et al.* 2009). To further confirm that the use of  $\beta$ -azide derivatives could be of general applicability for  $\alpha$ -glycosynthases,  $\beta$ -galactosyl-azides as possible donors for  $\alpha$ -galactosynthases, were considered. To this aim, the  $\alpha$ -galactosidase (TM1192) from the hyperthermophilic bacterium *Thermotoga maritima* (TmGalA) belonging to family GH36 was chosen as model system. Previously, this enzyme had been characterized and shown to follow the *retaining* reaction mechanism in which Asp327 and Asp387 are the nucleophile and the acid/base, respectively (Comfort *et al.* 2007). In addition, the three-dimensional (3D) structure at 2.3 Å resolution has also been deposited (pdb code 1ZY9). We show here that the mutant Asp327Gly is an efficient  $\alpha$ -galactosynthase producing different galactosylated disaccharides from  $\beta$ -galactosyl-azide donors and 4-nitrophenyl- $\alpha$ - and  $\beta$ -glycosides as acceptors. This is the first



$\alpha$ -galactosynthase produced so far and our results demonstrate that the use of  $\beta$ -azide derivatives is a general strategy for the preparation of  $\alpha$ -glycosynthases.

## Results and discussion

### *Kinetic characterization and chemical rescue of TmGalA nucleophile mutants*

The functional role of catalytic nucleophile in the  $\alpha$ -galactosidase from *T. maritima* (TmGalA) has been assigned to Asp327, based on structural alignment with other GH36 enzymes, analysis of the pH effect, azide rescue, and characterization of the products obtained with the mutant Asp327Gly (Comfort *et al.* 2007). These data prompted us to test if mutants in Asp327 could act as  $\alpha$ -galactosynthases. To this aim, we analyzed the mutants Asp327Ala and Gly, previously described (Comfort *et al.* 2007), and the newly prepared mutant Asp327Ser. The Asp327Ala and Ser mutant genes expressed in *Escherichia coli* were assayed at 65°C in 50 mM sodium acetate, pH 5.0, in permeabilized cells and were found to be completely inactive as expected. However, no  $\alpha$ -galactosidase activity could be rescued by adding 0.1–2.0 M sodium azide (not shown), suggesting that the two mutations had a detrimental effect on the activity and/or the stability of TmGalA. Therefore these mutants were not characterized any further.

It was previously reported that Asp327Gly mutant had its hydrolysis rate and catalytic efficiency reduced by 200–300-fold on 4-nitrophenyl- $\alpha$ -D-galactopyranoside (4NP- $\alpha$ -D-Gal) at 37°C if compared to wild type TmGalA (Comfort *et al.* 2007). This inactivation is lower than that commonly observed in other glycoside hydrolases mutated in the nucleophile (Viladot *et al.* 1998; Hrmova *et al.* 2002; Bravman *et al.* 2003; Zechel *et al.* 2003), but it is not uncommon among  $\alpha$ -glycosidases (Knegtel *et al.* 1995; Shallom *et al.* 2002; Tarling *et al.* 2003). However, when steady state kinetic constants were measured at 65°C, we observed a decrease of  $10^3$ - and  $2.6 \times 10^3$ -fold of the  $k_{cat}$  and  $k_{cat}/K_M$ , respectively, if compared to the wild

type (Table 1). Similar results were obtained in both sodium acetate and sodium citrate-phosphate buffers, indicating that acetate ions did not rescue the hydrolytic activity of the mutant and that this is the basal activity of Asp327Gly (Table 1).

The chemical rescue of the activity of the Asp327Gly in the presence of sodium azide was similar to that previously reported at 37°C (about 30-fold) (Table 1, Figure S1A), and the two enzymes showed a similar dependence on temperature of the  $\alpha$ -galactosidase activity (Figure S1B) (Comfort *et al.* 2007). Possibly, the inactivation produced by the mutation Asp327Gly was underestimated at 37°C because of the lower specific activity of the wild type at this temperature.

#### *Asp327Gly uses $\beta$ -galactosyl-azide as donor for $\alpha$ -galacto-oligosaccharide synthesis*

We have recently reported that two  $\alpha$ -L-fucosidases mutated in the nucleophile were able to promote  $\alpha$ -fucosynthetic reactions by using as donor substrate a fucosyl-azide derivative with an anomeric configuration opposite to that of the original substrate ( $\beta$ -fucosyl-azide) (Cobucci-Ponzano *et al.* 2009). By following the same approach, we tested the TmGalA Asp327Gly mutant for promoting transgalactosylation reactions from  $\beta$ -galactosyl-azide ( $\beta$ -Gal-N<sub>3</sub>), which was promptly chemically synthesized from galactose in three steps and 85% overall yield (see Material and methods). Incubations of the mutant at 65°C in 50 mM sodium acetate buffer, pH 5.0, in the presence of  $\beta$ -Gal-N<sub>3</sub> as unique substrate (at concentrations 1-14 mM) did not lead to any product by inspection of runs on TLC, indicating that no auto-condensation reactions were catalyzed at these conditions (not shown). Therefore, we included different glycosides in the reaction as possible acceptors at 1:1 and 1:10 donor: acceptor molar ratios. Most of the acceptors used did not lead to transgalactosylation products at these conditions, suggesting that TmGalA is rather selective in the -1 sub-site of the catalytic center, according to current nomenclature (Davies *et al.* 1997). However, UV-visible

products showing a different migration compared to substrates were clearly observed by TLC, when we used 4NP- $\alpha$ -D-Glc, -Man, -Xyl, and 4NP- $\beta$ -D-Xyl as acceptors at 1:1 molar ratio with the donor (Figure 2, Table S1). No transgalactosylation products could be observed when 4NP- $\alpha$ -D-Xyl and - $\beta$ -D-Xyl were used at 1:10 donor: acceptor molar ratios; possibly, at these conditions, they compete with  $\beta$ -Gal-N<sub>3</sub> in the +1 donor binding site.

To characterize the products, the reaction mixtures were scaled up and the transglycosylation products were purified, as described in the Materials and Methods section. In addition, the transgalactosylation efficiency was measured by running aliquots of the reaction by HPAEC-PAD. The results are summarized in Table II. In the presence of 4NP- $\alpha$ -Glc, only the compound **1** ( $\alpha$ -Gal-(1-6)- $\alpha$ -Glc-4NP) was synthesized in 33% yield (reaction I). Higher yields were obtained when 4NP- $\alpha$ -Xyl and 4NP- $\beta$ -Xyl were used as acceptors (40% and 38% for reactions II and III, respectively), obtaining, with the former, compounds **2** and **3** ( $\alpha$ -Gal-(1-2)- $\alpha$ -Xyl-4NP and  $\alpha$ -Gal-(1-4)- $\alpha$ -Xyl-4NP, respectively) and only compound **4** ( $\alpha$ -Gal-(1-4)- $\beta$ -Xyl-4NP) with 4NP- $\beta$ -Xyl acceptor. It is worth mentioning that the Asp327Gly mutant showed good regioselectivity with transfer of the galactose moiety exclusively onto a single OH. When 4NP- $\alpha$ -Glc was the acceptor, the enzyme transgalactosylated exclusively the primary alcohol at the C6. The other compounds observed by TLC (Table S1) were present in negligible amounts and could not be isolated. The regioselectivity on xyloside acceptors differs on the basis of the anomeric configuration with main products of transgalactosylation on the OH at the C2 and C4 groups of 4NP- $\alpha$ -Xyl and 4NP- $\beta$ -Xyl, respectively (Table 2). Possibly, this results from the different binding of these molecules on the -1 sub-site. In reaction IV, with 4NP- $\alpha$ -Man as acceptor, we could isolate two products, **5** and **6**, corresponding to  $\alpha$ -Gal-(1-6)- $\alpha$ -Man-4NP and  $\alpha$ -Gal-(1-3)- $\alpha$ -Man-4NP, respectively, with total yields of 51%. Again, the primary alcohol at the C6 was the preferred functional group, while the transgalactosylation product on the OH at the C3 was synthesized at very low

amounts as shown by the relative molar ratio of about 9:1 (Table 2). At these conditions, much donor substrate was left, but the total yields, although not quantitative as observed in other glycosynthases (Cobucci-Ponzano *et al.* 2009), are similar to those of  $\beta$ -glycosynthases from families GH16, GH26, GH52, and higher than those of the  $\alpha$ -fucosynthase from family GH95 (Fairweather *et al.* 2002; Jahn *et al.* 2003; Ben-David *et al.* 2007; Wada *et al.* 2008). In an effort to increase the efficiency of the transfer of  $\beta$ -Gal-N<sub>3</sub> on 4NP- $\alpha$ -Man (maintaining the same 1:1 molar ratio), we used 32-fold more  $\alpha$ -galactosynthase (320  $\mu$ g in 0.1 mL of reaction mixture): at these conditions, the conversion of the donor increased of 3-fold in a significantly shortened reaction (1 h) while the yields were 1.6-fold lower (data not shown). To compare our results to the transglycosylating activity of TmGalA, we incubated the wild type at 65°C in 50 mM sodium acetate buffer, pH 5.0, 14 mM  $\alpha$ -galactosyl-azide, and 14 mM 4NP- $\alpha$ -Man acceptor: no products could be observed (data not shown), demonstrating the utility of the galactosynthase approach.

It is worth mentioning that the production of  $\alpha$ -galactosides is generally a challenge in carbohydrate synthetic chemistry, as with 1,2-*cis*-glycosides no favorable effect, such as neighbouring group participation, is possible (Zhu and Schmidt 2009). Some chemical methods for 1,2-*cis*-galactosylations have been reported (Fukase *et al.* 1998; Demchenko *et al.* 1999; Imamura *et al.* 2003; Hada *et al.* 2006). However, these are generally based on the use of specific protecting group patterns, requiring tedious and yield-limiting chemical manipulations for their implementation, and typically with a lack of general utility.

This work represents the proof of principle that the method can be used for the synthesis of oligosaccharides of technological interest. Oligosaccharides containing terminal  $\alpha$ -Gal-(1-3)- $\beta$ -Gal sequence ( $\alpha$ -gal epitope), synthesized *in vivo* by the  $\alpha$ -1,3-galactosyltransferase ( $\alpha$ -1,3GT), have been identified as the major xenoactive antigen responsible for hyperacute rejection (Galili 2001). In addition, since anti-Gal is abundant and ubiquitous in humans, the

$\alpha$ -gal epitope has also clinical potential in the production of vaccines and in anticancer therapies (Macher and Galili 2008). Future work on the implementation of the  $\alpha$ -galactosynthase described here could open promising alternatives to  $\alpha$ -1,3GT for the production of  $\alpha$ -galacto-oligosaccharides.

#### *Reaction mechanism of $\alpha$ -D-galactosynthase*

We demonstrated that the TmGalA mutant Asp327Gly is a novel  $\alpha$ -galactosynthase, the first produced so far, to the best of our knowledge, and the third example of  $\alpha$ -glycosynthases along with  $\alpha$ -glucosynthase and  $\alpha$ -fucosynthase (Okuyama *et al.* 2002; Cobucci-Ponzano *et al.* 2009; Wada *et al.* 2008). The approach reported here, exploiting  $\beta$ -glycosyl-azides as donors, is the same recently proposed for two  $\alpha$ -fucosynthases from family GH29 (Cobucci-Ponzano *et al.* 2009). The mechanism followed by Asp327Gly  $\alpha$ -glycosynthase is shown in Figure 1D: the cavity created by the mutation that removed the side chain of Asp327 allowed the access to the active site of the  $\beta$ -Gal- $N_3$ . Then, the galactose moiety is transferred to the acceptor activated by base catalysis of the Asp387 residue; the disaccharide product, showing the newly formed  $\alpha$ -bond, cannot be hydrolyzed by the mutant and thus accumulates in the reaction mixture. This approach follows the same principle proposed for the first time by Withers and collaborators for  $\beta$ -glycosynthases exploiting fluoro-glycoside derivatives with an anomeric configuration opposite to that of the 'natural' substrate of the enzyme (Mackenzie *et al.* 1998). Here, following from GH29  $\alpha$ -fucosynthases (Cobucci-Ponzano *et al.* 2009), we confirm that  $\beta$ -azide-glycoside derivatives can be usefully exploited also by GH36  $\alpha$ -galactosynthases, demonstrating that azide derivatives are a valid alternative to fluoro derivatives for the production of  $\alpha$ -glycosynthases.

Glycosyl-azides are common donors in glycosidases-catalyzed synthesis and have been used for the screening of novel glycosidases mutants acting as thioglycoligases (Müllegger et al. 2005, Bojarová et al. 2007). In our case, glycosyl azides are crucial for the production of  $\alpha$ -glycosynthases, in fact,  $\beta$ -Fuc-N<sub>3</sub> and  $\beta$ -Gal-N<sub>3</sub> are more stable than their  $\beta$ -Fuc-F and  $\beta$ -Gal-F counterparts, because sodium azide is a less effective leaving group than fluoride. Therefore,  $\beta$ -glycosyl-azides are not degraded spontaneously at the operational conditions and, nonetheless, are sufficiently reactive donors. We demonstrated this property with the observation that  $\beta$ -fucosyl-,  $\beta$ -galactosyl-, and  $\beta$ -mannosyl-fluorides, being 6-deoxyhexopyranosides and/or showing axial substituents on the C2 and C4, are easily activated than hexopyranosides with C4 equatorial substituents, respectively (Overend 1972; Albert et al. 2000). It is likely that, the instability of  $\beta$ -fluoride derivatives have hampered the production of  $\alpha$ -glycosynthases; as an attractive alternative,  $\beta$ -azide derivatives can serve as substrates for the production of novel  $\alpha$ -glycosynthases from unrelated families of glycosidases.

## Materials and methods

### *Chemicals*

All commercially available substrates were purchased from Sigma-Aldrich. The galactosyl-azides were chemically synthesized from galactose according to a reported procedure (Györgydeák and Szilágyi 1987):  $\beta$ -Gal- $N_3$  in three steps (peracetylation, stereoselective anomeric azidation and deacetylation) and 85% overall yield while  $\alpha$ -Gal- $N_3$  in five steps (peracetylation, anomeric  $\alpha$ -bromination,  $\beta$ -chlorination, azidation and deacetylation) and 16% yield.

### *Site-directed mutagenesis, expression and purification of wild type and mutant TmGalA*

The plasmids expressing TmGalA wild type and Asp327Gly and Ala mutants were described previously (Comfort *et al.* 2007). The mutant Asp327Ser was generated using the GeneTailor Site-directed Mutagenesis System (Invitrogen) and the synthetic oligonucleotides (PRIMM, Italy) in which the mutated nucleotides are underlined:

TMSerfwd:

5'-GCTACAGGTA~~CTT~~CAAGATCTTCTTTCTCTTCG-3'

TMSerrev:

5'-GATCTTGAAGTACCTGTAGCCCATCTTTCT-3'

The gene containing the desired mutation was identified by direct sequencing and completely re-sequenced.

The wild type and mutants  $\alpha$ -galactosidases from *T. maritima* were expressed in *E. coli* BL21(DE3) and purified as previously described (Comfort *et al.* 2007). After the last purification step, the enzymes were extensively dialyzed against 20 mM sodium phosphate pH 7.3; 150 mM NaCl.

### *Enzymatic characterization*

The  $\alpha$ -galactosidase activity assays, the steady state kinetic studies, and the azide rescue reactions were performed as previously reported (1  $\mu$ g of wild type and 10  $\mu$ g of Asp327Gly mutant), but at 65°C and without added bovine serum albumin (Comfort *et al.* 2007). Suitable blanks, containing all the reagents with the exception of enzyme, were always used to take into account the negligible spontaneous hydrolysis of the substrates. One enzymatic unit is defined as the amount of enzyme catalyzing the conversion of one  $\mu$ mole of substrate into product in one min, at the indicated conditions. All kinetic data were calculated as the average of at least two experiments and were plotted and refined with the program *GraFit* (Leatherbarrow 1992).

The effect of temperature on the activity was analyzed by measuring the  $\alpha$ -galactosidase activity of wild type and Asp327Gly mutant (1  $\mu$ g and 10  $\mu$ g, respectively) in 50 mM sodium acetate buffer, pH 5.0, in the temperature range 37-80°C. Blanks with no enzyme were used at each temperature to subtract the spontaneous hydrolysis of the substrate.

Assays of TmGalA wild type and mutants on permeabilized cells of *E. coli* BL21(DE3) were performed by the method of Zhang and Bremer at 65°C in 50 mM sodium acetate buffer, pH 5.0, 2.5 mM 4NP- $\alpha$ -Gal, and 0.1-2.0 M sodium azide (Zhang and Bremer 1995).

### *Transgalactosylation trials*

The glycosynthetic reactions were performed by incubating Asp327Gly (10  $\mu$ g) for 16 h at 65°C in 0.1 mL of 50 mM sodium acetate buffer pH 5.0 at the indicated concentrations of  $\beta$ -Gal-N<sub>3</sub> (donor) and the suitable acceptor. Blank mixtures without enzyme were always prepared. The products distribution was evaluated by thin layer chromatography (TLC) as previously reported (Cobucci-Ponzano *et al.* 2009). The transgalactosylation efficiency of Asp327Gly mutant was measured by use of a High-Performance Anion-Exchange



Chromatography with Pulsed Amperometric Detection (HPAEC–PAD) equipped with a PA200 column (Dionex, USA). Samples were eluted with 20 mM NaOH at a flowrate of 0.5 ml min<sup>-1</sup> and the data were analysed as previously reported (Cobucci-Ponzano *et al.* 2009). In particular, to measure the total amount of galactose enzymatically transferred (total transgalactosylation efficiency shown in Table II), 1/10 of the reaction mixtures were incubated for 90 min at 65°C in the presence of 5.2 µg TmGalA wild type. The efficiency of the transgalactosylation reaction was calculated as: total amount of galactose transferred - moles of galactose transferred to water / total amount of galactose transferred x 100.

#### *Characterization of the galactosylated oligosaccharides*

The glycosynthetic reactions were performed at the same conditions described above in a total volume of 2 ml. The transgalactosylation products were purified by reverse phase chromatography (Polar-RP 80A, Phenomenex, 4 µ, 250 x 10 mm) on an Agilent HPLC instrument 1100 series and revealed by UV at 220 nm. Samples were eluted using the following conditions: 40% of methanol in water for reaction I; 40% of methanol in water for 5 min, 40% to 50% in 30 min, 50% for 15 min for reaction II; 30% methanol in water for 5 min, 30% to 50% in 30 min, 50% for 15 min for reactions III and IV. For all the disaccharides isolated except for the product **3**, the structural determination was obtained by <sup>1</sup>H mono-dimensional and homonuclear (<sup>1</sup>H, <sup>1</sup>H) and heteronuclear (<sup>1</sup>H, <sup>13</sup>C) two-dimensional NMR experiments and by methylation analysis.

#### **Product 1:**

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 308K): δ 5.87 (d, 1H, *J*<sub>H-1,H-2</sub>=3.6 Hz, H-1A), 4.88 (d, 1H, *J*<sub>H-1,H-2</sub>=3.7 Hz, H-1B), 3.95 (t, 1H, H-3A), 3.93 (d, 1H, H-4B), 3.90 (t, 1H, H-5A), 3.87 (d, 1H, H-6aA), 3.87 (d, 1H, H-5B), 3.82 (dd, 1H, H-2A), 3.75 (d, 1H, H-6aB), 3.74 (d, 1H, H-2B), 3.72 (d, 1H, H-6bB), 3.71 (d, 1H, H-6bA), 3.58 (dd, 1H, H-3B), 3.56 (t, 1H, H-4A); <sup>13</sup>C NMR (150

MHz, D<sub>2</sub>O):  $\delta$  99.2 (C-1B), 97.7 (C-1A), 74.5 (C-3A), 73.2 (C-5B), 72.2 (C-2A), 72.2 (C-5A), 70.8 (C-4A), 70.8 (C-3B), 70.5 (C-4B), 69.6 (C-2B), 66.9 (C-6A), 62.4 (C-6B).

**Product 2:**

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 298K):  $\delta$  5.93 (d, 1H,  $J_{H-1,H-2}$ =3.3 Hz, H-1A), 5.00 (d, 1H,  $J_{H-1,H-2}$ =3.8 Hz, H-1B), 4.09 (t, 1H, H-5B), 3.90 (d, 1H, H-3A), 3.90 (d, 1H, H-4B), 3.82 (dd, 1H, H-3B), 3.79 (dd, 1H, H-2A), 3.69 (t, 1H, H-5aA), 3.68 (d, 1H, H-2B), 3.67 (m, 1H, H-4A), 3.64 (d, 2H, H-6a,bB), 3.48 (t, 1H, H-5bA); <sup>13</sup>C NMR (150 MHz, D<sub>2</sub>O):  $\delta$  97.5 (C-1B), 95.3 (C-1A), 76.0 (C-2A), 72.7 (C-3A), 72.3 (C-5B), 70.3 (C-4B), 70.2 (C-3B), 70.0 (C-4A), 69.3 (C-2B), 63.3 (C-5A), 62.2 (C-6B).

**Product 4:**

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 298K):  $\delta$  5.12 (d, 1H,  $J_{H-1,H-2}$ =7.7 Hz, H-1A), 5.10 (d, 1H,  $J_{H-1,H-2}$ =3.7 Hz, H-1B), 4.20 (dd, 1H, H-5aA), 3.87 (dd, 1H, H-3B), 3.86 (t, 1H, H-5B), 3.75 (d, 1H, H-4B), 3.74 (dd, 1H, H-2B), 3.69 (t, 1H, H-3A), 3.68 (dd, 1H, H-4A), 3.65 (dd, 2H, H-6a,bB), 3.56 (t, 1H, H-2A), 3.56 (m, 1H, H-5bA); <sup>13</sup>C NMR (150 MHz, D<sub>2</sub>O):  $\delta$  101.5 (C-1B), 101.0 (C-1A), 78.7 (C-4A), 75.5 (C-3A), 73.6 (C-2A), 72.7 (C-5B), 70.5 (C-3B), 70.4 (C-4B), 69.7 (C-2B), 65.6 (C-5A), 62.4 (C-6B).

**Product 5:**

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 315K):  $\delta$  5.87 (bs, 1H, H-1A), 4.95 (d, 1H,  $J_{H-1,H-2}$ =3.8 Hz, H-1B), 4.30 (dd, 1H, H-2A), 4.14 (dd, 1H, H-3A), 4.00 (d, 1H, H-4B), 3.98 (t, 1H, H-5A), 3.97 (dd, 1H, H-6aA), 3.90 (t, 1H, H-5B), 3.89 (t, 1H, H-4A), 3.82 (dd, 1H, H-6aB), 3.81 (dd, 1H, H-2B), 3.78 (dd, 1H, H-6bB), 3.74 (dd, 1H, H-6bA), 3.66 (dd, 1H, H-3B); <sup>13</sup>C NMR (150 MHz, D<sub>2</sub>O):  $\delta$  99.3 (C-1B), 99.2 (C-1A), 73.8 (C-5B), 72.3 (C-5A), 72.0 (C-3A), 71.1 (C-2A), 71.0 (C-3B), 70.6 (C-4B), 69.8 (C-2B), 68.0 (C-4A), 67.0 (C-6A), 62.4 (C-6B).

**Product 6:**

$^1\text{H}$  NMR (600 MHz,  $\text{D}_2\text{O}$ , 298K):  $\delta$  5.66 (bs, 1H, H-1A), 5.23 (d, 1H,  $J_{\text{H-1},\text{H-2}}=4.0$  Hz, H-1B), 4.28 (dd, 1H, H-2A), 4.10 (dd, 1H, H-3A), 4.00 (dd, 1H, H-5B), 3.90 (d, 1H, H-4B), 3.85 (dd, 1H, H-3B), 3.84 (dd, 1H, H-4A), 3.73 (dd, 1H, H-2B), 3.63 (dd, 1H, H-6aA), 3.65 (dd, 1H, H-6aB), 3.67 (dd, 1H, H-6bB), 3.68 (dd, 1H, H-6bA), 3.56 (dd, 1H, H-5A);  $^{13}\text{C}$  NMR (150 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  101.9 (C-1B), 98.9 (C-1A), 79.3 (C-3A), 75.0 (C-5A), 72.7 (C-5B), 70.7 (C-2A), 70.7 (C-3B), 70.5 (C-4B), 69.5 (C-2B), 66.6 (C-4A), 62.2 (C-6B), 61.7 (C-6A).

The disaccharide **3** was isolated in a very low amount preventing us from characterizing it by two-dimensional NMR experiments. Therefore the 4NP- $\alpha$ -D-Xyl substitution was identified by methylation analysis which revealed the presence of terminal-galactose and 4-substituted xylose, by comparison with authentic standards. The  $^1\text{H}$ -NMR experiment revealed for both the xylose (H-1A,  $\delta$  5.7) and the galactose units (H-1B,  $\delta$  5.1) an  $\alpha$  configuration ( $J_{\text{H-1A},\text{H-2A}}=3.6$  Hz,  $J_{\text{H-1B},\text{H-2B}}=3.1$  Hz).

The relative ratio reported in Table II was obtained integrating the UV peaks area in the HPLC chromatogram.

## Funding

This work was supported by the Agenzia Spaziale Italiana [project MoMa n. 1/014/06/0].

R.M.K. acknowledges support from the US National Science Foundation.

## Abbreviations

CAZy: Carbohydrate Active enZYme classification; Gal: galactopyranoside;  $\beta$ -Gal-N<sub>3</sub>:  $\beta$ -galactosyl-azide; GH: glycoside hydrolase; Glc: glucopyranoside; GT glycosyltransferases; HPAEC–PAD: High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection; 2- or 4-NP: 2- or 4-nitrophenyl; Man: mannopyranoside; TLC: thin layer chromatography; TmGalA:  $\alpha$ -galactosidase from *Thermotoga maritima*; Xyl: xylopyranoside

## References

- Albert M, Repetschnigg W, Ortner J, Gomes J, Paul BJ, Illaszewicz C, Weber H, Steiner W, Dax K. 2000. Simultaneous detection of different glycosidase activities by F-19 NMR spectroscopy. *Carbohydr Res* 327, 395-400.
- Ben-David A, Bravman T, Balazs YS, Czjzek M, Schomburg D, Shoham G, Shoham Y. 2007. Glycosynthase activity of *Geobacillus stearothermophilus* GH52 beta-xylosidase: efficient synthesis of xylooligosaccharides from alpha-D-xylopyranosyl fluoride through a conjugated reaction. *ChemBiochem* 8, 2145-2151.
- Bojarová P, Petrásková L, Ferrandi EE, Monti D, Pelantová H, Kuzma M, Simerská P and Křen V. 2007. Glycosyl Azides – An Alternative Way to Disaccharides. *Adv. Synth. Catal.* 349, 1514-1520.
- Bravman T, Belakhov V, Solomon D, Shoham G, Henrissat B, Baasov T, Shoham Y. 2003. Identification of the catalytic residues in family 52 glycoside hydrolase, a beta-xylosidase from *Geobacillus stearothermophilus* T-6. *J Biol Chem* 278, 26742-26749.
- Brown JR, Crawford BE, Esko JD 2008. Glycan antagonists and inhibitors: a fount for drug discovery. *Crit Rev Biochem Mol.* 42, 481-515.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 37, D233-D238.
- Cobucci-Ponzano B, Conte F, Bedini E, Corsaro MM, Parrilli M, Sulzenbacher G, Lipski A, Dal Piaz F, Lepore L, Rossi M, Moracci M. 2009.  $\beta$ -Glycosyl azides as substrates for  $\alpha$ -glycosynthases: preparation of efficient  $\alpha$ -L-fucosynthases. *Chem Biol* 16, 1097-1108.
- Comfor, DA, Bobrov KS, Ivanen DR, Shabalin KA, Harris JM, Kulminskaya AA, Brumer H, Kelly RM 2007. Biochemical analysis of *Thermotoga maritima* GH36  $\alpha$ -galactosidase

- (TmGalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases. *Biochemistry* 46, 3319-3330.
- Davies GJ, Wilson KS, Henrissat B. 1997. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* 321, 557-559.
- Demchenko AV, Rousson E, Boons GJ 1999. Stereoselective 1,2-cis-galactosylation assisted by remote neighboring group participation and solvent effects. *Tetrahedron Lett.* 40, 6523-6526.
- Drake PM, Cho W, Li BS, Prakobphol A, Johansen E, Anderson NL, Regnier FE, Gibson BW, Fisher SJ 2010. Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin Chem* 56, 223-236.
- Fairweather JK, Faijes M, Driguez H, Planas A. 2002. Specificity studies of *Bacillus* 1,3-1,4-beta-glucanases and application to glycosynthase-catalyzed transglycosylation. *ChemBiochem* 3, 866-873.
- Fukase K, Nakai Y, Kanoh T, Kusumoto S. 1998. Mild but efficient methods for stereoselective glycosylation with thioglycosides: activation by [N-phenylselenophthalimide-Mg(ClO<sub>4</sub>)<sub>2</sub>] and [PhIO-Mg(ClO<sub>4</sub>)<sub>2</sub>]. *Synlett* 84-86.
- Gabius HJ, Siebert HC, Andre S, Jimenez-Barbero J, Rudiger H. 2004. Chemical biology of the sugar code. *ChemBiochem* 5, 741-764.
- Galili U. 2001. The  $\alpha$ -gal epitope (Gal  $\alpha$  1-3Gal  $\beta$  1-4GlcNAc-R) in xenotransplantation. *Biochimie* 83, 557-563.
- Györgydeák ZN, Szilágyi LS. 1987. Einfache synthesen der anomeren, an C-6 modifizierten galacto- und glucopyranosylazide. *Liebigs Ann.* 1987, 235-241.
- Hada N, Oka J, Nishiyama A, Takeda T. 2006. Stereoselective synthesis of 1,2-cis galactosides: synthesis of a glycolipid containing Gal  $\alpha$  1-6Gal component from *Zygomycetes* species. *Tetrahedron Lett* 47, 6647-6650.

- Hancock SM, Vaughan M, Withers SG. 2006. Engineering of glycosidases and glycosyltransferases. *Curr Opin Chem Biol* 10, 509-519.
- Honda Y, Kitaoka M. 2006. The first glycosynthase derived from an inverting glycoside hydrolase. *J Biol Chem* 281, 1426-1431.
- Hrmova M, Imai T, Rutten SJ, Fairweather JK, Pelosi L, Bulone V, Driguez H, Fincher GB. 2002. Mutated barley (1,3)- $\beta$ -D-glucan endohydrolases synthesize crystalline (1,3)- $\beta$ -D-glucans. *J Biol Chem* 277, 30102-30111.
- Imamura A, Ando H, Korogi S, Tanabe G, Muraoka O, Ishida H, Kiso M. 2003. Di-tert-butylsilylene (DTBS) group-directed  $\alpha$ -selective galactosylation unaffected by C-2 participating functionalities. *Tetrahedron Lett* 44, 6725-6728.
- Jahn M, Stoll D, Warren RAJ, Szabo L, Singh P, Gilbert HJ, Ducros VMA, Davies GJ, Withers SG. 2003. Expansion of the glycosynthase repertoire to produce defined manno-oligosaccharides. *Chem Commun* 1327-1329.
- Kiessling LL, Splain RA. 2010. Chemical approaches to glycobiology. *Ann Rev Biochem* 79, 619-653.
- Knegtel RMK, Strokopytov B, Penninga D, Faber OG, Rozeboom HJ, Kalk KH, Dijkhuizen L, Dijkstra BW. 1995. Crystallographic studies of the interaction of cyclodextrin glycosyltransferase from *Bacillus circulans* strain-251 with natural substrates and products. *J Biol Chem* 270, 29256-29264.
- Koshland DE. 1953. Stereochemistry and the mechanism of enzymatic reactions. *Biol Rev* 28, 416-436.
- Leatherbarrow RJ. 1992. GraFit. (Erithacus Software Ltd: Staines, U.K).
- Macher BA, Galili U. 2008. The Gal  $\alpha$  1,3Gal  $\beta$  1,4GlcNAc-R ( $\alpha$ -Gal) epitope: a carbohydrate of unique evolution and clinical relevance. *BBA Gen Subjects* 1780, 75-88.

- Mackenzie LF, Wang QP, Warren RAJ, Withers SG. 1998. Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J Am Chem Soc* 120, 5583-5584.
- Malet C, Planas A. 1998. From  $\beta$ -glucanase to beta-glucansynthase: glycosyl transfer to  $\alpha$ -glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS Lett* 440, 208-212.
- McCarter JD, Withers SG. 1994. Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struc Biol* 4, 885-892.
- Moracci M, Trincone A, Perugino G, Ciamarella M, Rossi M. 1998. Restoration of the activity of active-site mutants of the hyperthermophilic  $\beta$ -glycosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophiles. *Biochemistry* 37, 17262-17270.
- Müllegger J, Jahn M, Chen HM, Warren RA, Withers SG. 2005. Engineering of a thioglycoligase: randomized mutagenesis of the acid-base residue leads to the identification of improved catalysts. *Protein Eng Des Sel* 18, 33-40.
- Okuyama M, Mori H, Watanabe K, Kimura A, Chiba S. 2002.  $\alpha$ -glucosidase mutant catalyzes " $\alpha$ -glycosynthase"-type reaction. *Biosci Biotech Bioch* 66, 928-933.
- Overend WG. 1972. *The carbohydrates* (New York and London Academic Press).
- Perugino G, Trincone A, Rossi M, Moracci M. 2004. Oligosaccharide synthesis by glycosynthases. *Trends Biotechnol* 22, 31-37.
- Seeberger PH. 2008. Automated oligosaccharide synthesis. *Chem Soc Rev* 37, 19-28.
- Shaikh FA., Withers SG. 2008. Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis. *Biochem Cell Biol* 86, 169-177.



- Shallom D, Belakhov V, Solomon D, Shoham G, Baasov T, Shoham Y. 2002. Detailed kinetic analysis and identification of the nucleophile in  $\alpha$ -L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *J Biol Chem* 277, 51084-51084.
- Tarling CA, He SM, Sulzenbacher G, Bignon C, Bourne Y, Henrissat B, Withers SG. 2003. Identification of the catalytic nucleophile of the family 29  $\alpha$ -L-fucosidase from *Thermotoga maritima* through trapping of a covalent glycosyl-enzyme intermediate and mutagenesis. *J Biol Chem* 278, 47394-47399.
- Vanhooren PT, Vandamme EJ. 1999. L-fucose: occurrence, physiological role, chemical, enzymatic and microbial synthesis. *J Chem Technol Biot* 74, 479-497.
- Viladot JL, de Ramon E, Durany O, Planas A. 1998. Probing the mechanism of *Bacillus* 1,3-1,4- $\beta$ -D-glucan 4-glucanohydrolases by chemical rescue of inactive mutants at catalytically essential residues. *Biochemistry* 37, 11332-11342.
- Wada J, Honda Y, Nagae M, Kato R, Wakatsuki S, Katayama T, Taniguchi H, Kumagai H, Kitaoka M, Yamamoto K. 2008. 1,2- $\alpha$ -L-Fucosynthase: a glycosynthase derived from an inverting  $\alpha$ -glycosidase with an unusual reaction mechanism. *FEBS Lett* 582, 3739-3743.
- Zechel DL, Reid SP, Stoll D, Nashiru O, Warren RAJ, Withers SG. 2003. Mechanism, mutagenesis, and chemical rescue of beta-mannosidase from *Cellulomonas fimi*. *Biochemistry* 42, 7195-7204.
- Zhang XG, Bremer H. 1995. Control of the *Escherichia coli* rrnB P1 promoter strength by ppGpp. *J Biol Chem* 270, 11181-11189.
- Zhu X, Schmidt RR. 2009. New principles for glycoside-bond formation. *Angew Chem Int Ed*. 48, 1900-1934.

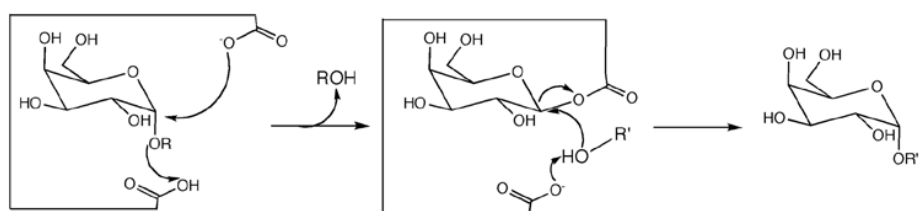
## Figure legends

Figure 1: *Reaction mechanism of glycosidases and glycosynthases* – (A) Reaction mechanism of a retaining  $\alpha$ -galactosidase; R = leaving group; R'OH = H<sub>2</sub>O: hydrolysis; R'OH = alcohol: transgalactosylation. (B) Reaction mechanisms of  $\beta$ -glycosynthases; (C) reaction mechanism of  $\alpha$ -fucosynthases. (D) Reaction mechanism of  $\alpha$ -galactosynthases.

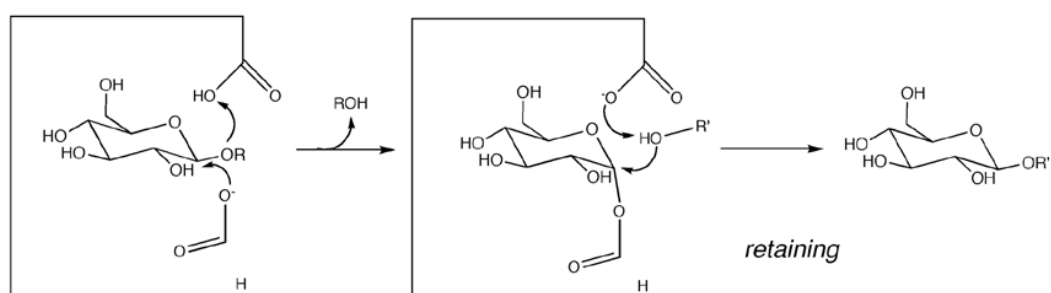
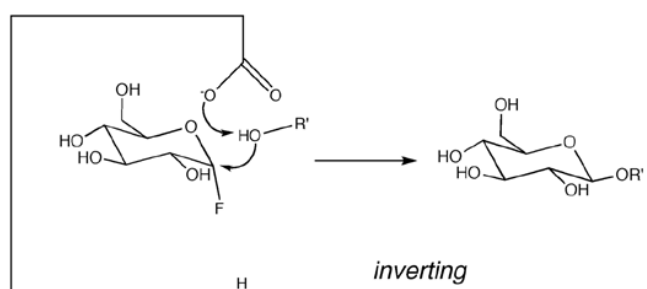
Figure 2: *TLC analysis of the transgalactosylation reactions of the Asp327Gly mutant* – (A) Reactions with 4NP- $\alpha$ -Glc and 4NP- $\alpha$ -Man acceptors. Lane 1: Gal standard (20  $\mu$ l 100 mM); lane 2:  $\beta$ -Gal-N<sub>3</sub> standard (20  $\mu$ l 80 mM); lane 3: reaction containing Asp327Gly (20  $\mu$ g),  $\beta$ -Gal-N<sub>3</sub> (14 mM), 4NP- $\alpha$ -Glc (14 mM), in 50 mM sodium acetate buffer, pH 5.0, at 65°C for 16 hrs; lane 4: blank control (same as lane 3, but without enzyme); lane 5: standard 4NP- $\alpha$ -Glc (20  $\mu$ l 100 mM); lane 6: Glc standard (20  $\mu$ l 100 mM); lane 7: reaction containing Asp327Gly (20  $\mu$ g),  $\beta$ -Gal-N<sub>3</sub> (14 mM), 4NP- $\alpha$ -Man, (14 mM), in 50 mM sodium acetate buffer, pH 5.0, at 65°C for 16 hrs; lane 8: blank control (same as lane 7, but without enzyme); lane 9: standard 4NP- $\alpha$ -Man (20  $\mu$ l 100 mM); lane 10: Man standard (20  $\mu$ l 100 mM). (B) Reactions with 4NP- $\alpha$ -Xyl and 4NP- $\beta$ -Xyl acceptors. Lane 1:  $\beta$ -Gal-N<sub>3</sub> standard (20  $\mu$ l 80 mM); lane 2: reaction containing Asp327Gly (20  $\mu$ g),  $\beta$ -Gal-N<sub>3</sub> (14 mM), 4NP- $\alpha$ -Xyl (14 mM) in 50 mM sodium acetate buffer, pH 5.0, at 65°C for 16 hrs; lane 3: blank control (same as lane 2, but without enzyme); lane 4: standard 4NP- $\alpha$ -Xyl (20  $\mu$ l 100 mM); lane 5: Xyl standard (20  $\mu$ l 100 mM); lane 6: reaction containing Asp327Gly (20  $\mu$ g),  $\beta$ -Gal-N<sub>3</sub> (14 mM), 4NP- $\beta$ -Xyl, (14 mM) in 50 mM sodium acetate buffer, pH 5.0, at 65°C for 16 hrs; lane 7: blank control (same as lane 6, but without enzyme); lane 8: standard 4NP- $\beta$ -Xyl (20  $\mu$ l 100

mM); lane 9: Gal standard (20  $\mu$ l 100 mM). The products were separated on a silica gel 60 F<sub>254</sub> TLC using ethyl acetate-methanol-water (70:20:10) as eluent and were detected by exposure to 4%  $\alpha$ -naphthol in 10% sulfuric acid in ethanol followed by charring. UV-visible reaction products are highlighted with ovals.

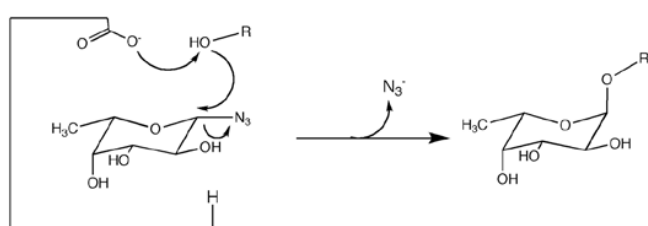
A



B



C



D

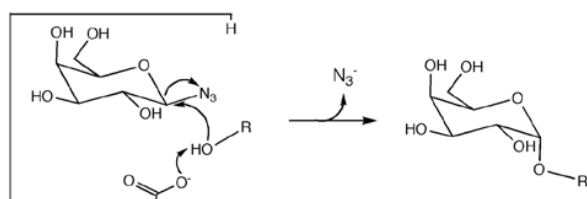


Figure 1

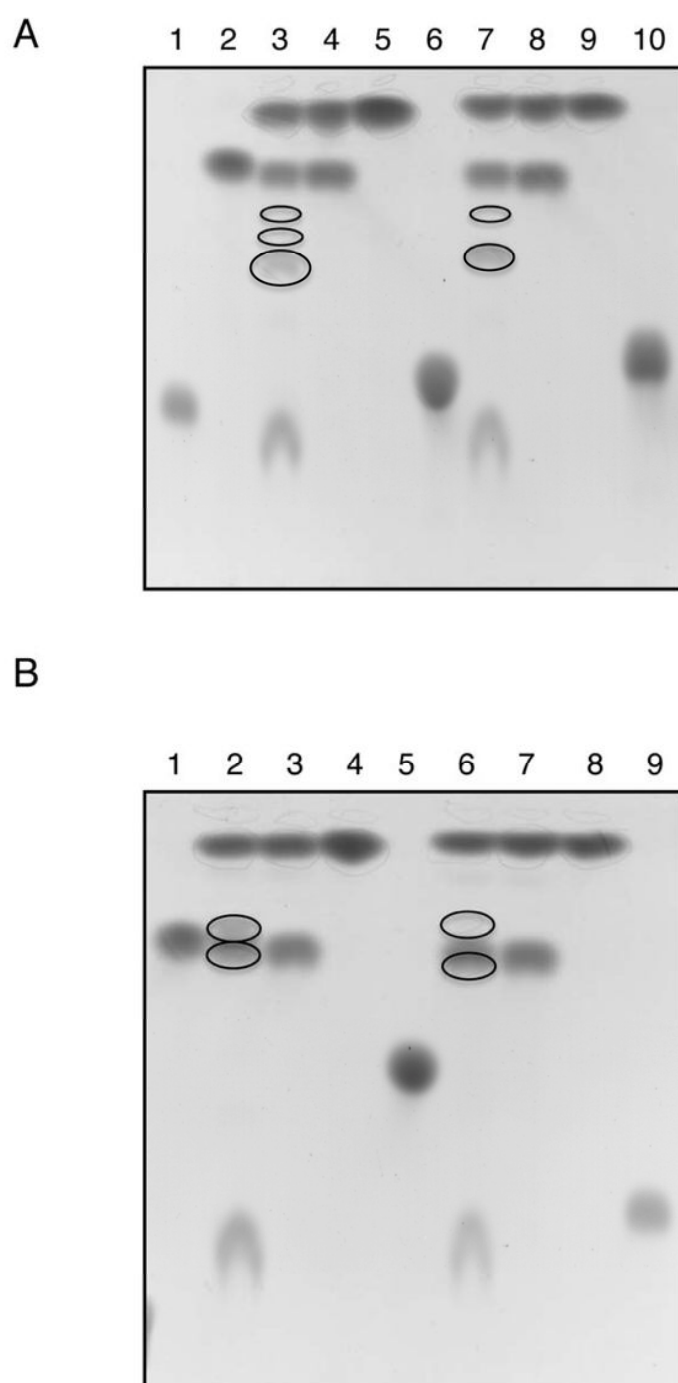
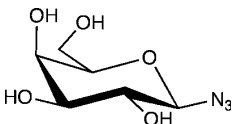
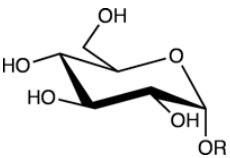
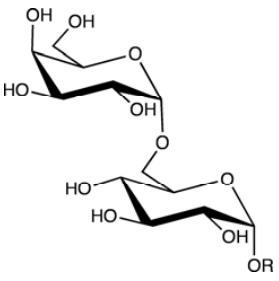
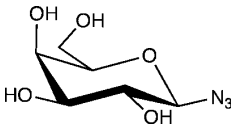
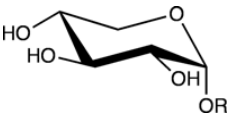
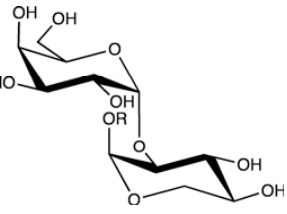


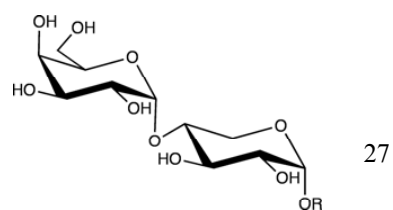
Figure 2

**Table I****Steady-state kinetic constants of wild type and mutant TmGalA**

	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ (mM)	$k_{cat}/K_M$ (s <sup>-1</sup> mM <sup>-1</sup> )
Wild-type	65±2	0.04±0.02	1585
D327G	0.060±0.003	0.10±0.05	0.6
D327G <sup>1</sup>	0.060±0.003	0.10±0.04	0.6
D327G in 500 mM NaN <sub>3</sub>	2.07±0.04	0.12±0.01	17
Assays were performed in 50 mM sodium acetate, pH 5.0 at 65°C on 2.5 mM 4NP-α-D-Gal			
<sup>1</sup> Assays were performed at the same conditions but in 50 mM sodium citrate-phosphate, pH 5.0			

**Table II**  
**Synthetic products of TmGalAD327G**

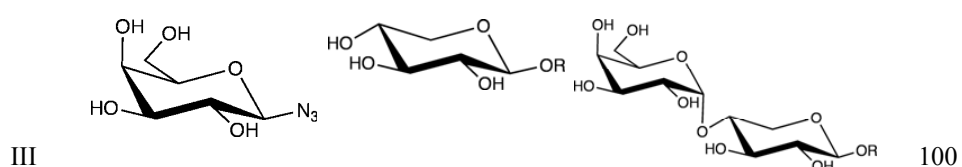
Reaction	Donor	Acceptor	Products	Relative ratios (%)
I			 <i>1</i>	100
Total transgalactosylation efficiency 33%				
II			 <i>2</i>	73



3

Total transgalactosylation

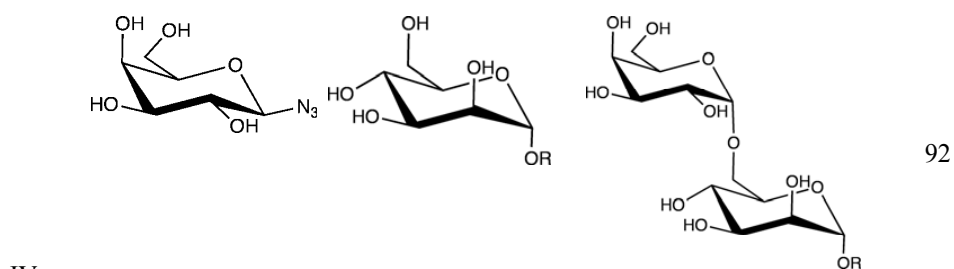
efficiency 40%



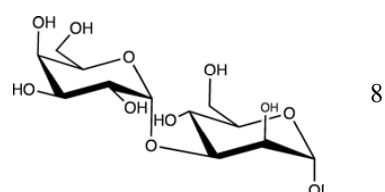
4

Total transgalactosylation

efficiency 38%



5



8



**6**

Total transgalactosylation  
efficiency 51%

---

R = 4-nitrophenol

---

## **Ringraziamenti**

Il lavoro qui descritto è il frutto di tre anni di attività svolta presso l'Istituto di Biochimica delle Proteine del C.N.R, anni che reputo sicuramente tra i più formativi della mia vita, grazie soprattutto alle magnifiche persone che hanno condiviso con me la propria conoscenza ed esperienza.

Ed è per questo che voglio ringraziare innanzitutto il mio tutor, Prof. Mosè Rossi che, come Direttore dell'IBP, mi ha dato la possibilità di lavorare nel suo Istituto e che mi ha seguito durante l'intero dottorato. Un ringraziamento particolare va al Dott. Marco Moracci, la persona che più di tutte mi ha seguito e supportato lungo questo mio percorso offrendomi ogni giorno la possibilità di crescere scientificamente e professionalmente.

Ringrazio inoltre la Dott.ssa Beatrice Cobucci-Ponzano per la sua amicizia, i suoi preziosi consigli e soprattutto per avermi ospitato, con pazienza, nel suo studio condividendo con me l'entropia della sua scrivania. Ringrazio Mila e Carmen per avermi supportato in questi anni, per la loro squisita compagnia per avermi offerto sempre un sorriso, anche quando non lo meritavo.

Infine ringrazio di cuore i miei genitori che dal primo momento mi hanno spinto a intraprendere questo percorso supportandomi moralmente, e non solo, durante questi tre anni, così come Roberta che mi è stata sempre vicina e mi ha spronato a fare sempre più e meglio.